Zorginstituut Nederland

Guideline for economic evaluations in healthcare

Datum   16 juni 2016
Status   Definitief

Colofon

Volgnummer	2016077622
Contactpersoon	mw. S. Knies +31 (0)20 797 87 51
Afdeling	Pakket
Uitgebracht aan	

Inhoud

Preface

The healthcare sector is a major focus of interest currently, and aspects such as quality, safety and effectiveness have become core concepts within this context. Increasingly, questions such as: 'What constitutes effective healthcare?' and What care should be reimbursed by the basic health insurance package?' are carefully being considered. This can only be done properly on the basis of data assessed by standard methods.

The present guideline document, entitled "Guideline for economic evaluations in healthcare", represents an update of previous guidelines from 2006 (Guidelines for pharmacoeconomic research), 2008 (Guidance on outcome research) and 2010 (Cost research manual). The document is comprised of the actual text of the guideline dedicated to the different elements of a health economic evaluation and additionally, a number of in-depth modules in which specific elements are addressed in more detail. In parallel to this revision, the much used Cost research manual was updated as well and is included as one of the in-depth modules in this document.

The committee charged with the guideline revision has done the job relatively quickly. Four meetings were needed, as well as the active commitment of the staff of the National Health Care Institute and the committee members, to update the 2006 guidelines. The committee took advantage of experiences gained during the previous guideline development, new insights such as social cost-benefit analysis, and of various publications by international organizations (for example EUnetHTA) and professional associations such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

The Guideline as it stands provides a well-updated overview of the methods used and recommended by the National Health Care Institute in the Netherlands. The basic approach to a cost-effectiveness study is summarized in the 'reference case'. Although the reference case should be strictly followed, one may deviate from certain elements of the Guideline provided sufficient grounds to do so. That is, in particular when dealing with healthcare products other than pharmaceuticals (such as medical devices), types of care other than curative care (such as long-term care) or when used in for example healthcare institutions.

Health-economic research for the purpose of reimbursement decisions is still an evolving field. To illustrate this, an international debate is ongoing about the use of multi-criteria decision analysis and patient preference studies, increasing knowledge is being gained about the use of real-world data, the issue of pharmacoeconomic research in personalized or precision medicine is being addressed, and there is an increasing interest in the application of economic evaluations in the developmental stage of products, which is known as early health technology assessment (HTA). All these developments should be monitored closely and be considered for inclusion in future versions of this Guideline.

On behalf of the committee charged with the guideline revision,

Prof. Maarten J. IJzerman
Chair

Glossary

Term	Description
Assumptions	Starting points for the analysis
BIA	Budget impact analysis, estimation of the total financial consequences of the use of a new intervention
CBA	Cost-benefit analysis, analysis of the difference in costs relative to the difference in benefits, with the benefits expressed in monetary units
CEA	Cost-effectiveness analysis, analysis of the difference in costs relative to the difference in (clinical) effects
CEAC	Cost-effectiveness acceptability curve, a graph which for a series of thresholds shows the probability that the intervention is cost-effective at the threshold in question
CUA	Cost-utility analysis, analysis of the difference in costs relative to the difference in QALYs
ICER	Incremental cost-effectiveness ratio, ratio of incremental costs to incremental effects of an intervention and the standard intervention
ICB	Intersectoral costs and benefits, intersectoral costs and benefits in sectors outside the healthcare, notably relevant in preventive interventions
Incremental	The difference between the new intervention and the standard and/-or usual treatment
LYG	Life years gained, number of years gained due to intervention
Perspective	Point of view from which the costs and benefits are considered
PSA	'Probabilistic sensitivity analysis', sensitivity analysis of the degree of uncertainty in the ICER as a result of the uncertainties around input parameters of a decision model
QALYs	Quality adjusted life years, the number of gained life years corrected for quality of life
Sensitivity analysis	Analysis to test the impact of uncertainty around input parameters, usually of a decision model, on the ICER
Univariate	When only one input parameter in the decision model is varied
Utility	The term 'utility' is often also used to refer to the quality of weighted lives in the QALY-analysis
VOI	Value of information, value of additional information to reduce the uncertainty around a decision

Summary

In 1999 the Health Care Insurance Board (College voor Zorgverzekeringen, CVZ) issued the 'Guidelines for pharmacoeconomic research'. An updated version (2005) was adopted in April 2006. This document provided an analytic framework for the pharmacoeconomic evaluations which are part of a reimbursement file. The framework serves to assess whether there is sufficient evidence for the efficacy of a drug for which reimbursement has been requested. Previously, three sets of guidelines were available, namely the above-mentioned Guidelines for pharmacoeconomic research (2006), the Guidance on outcome research (2008) and the Cost research manual (2010).

These three documents had been published under the name and under the auspices of CVZ. The very fact that they were published in different years was one of the reasons that the various recommendations did not always tally. This situation does not promote clarity. Furthermore, being outdated they do no longer conform to the most recent developments in the (international) research field. Lastly, those guidelines were not always directly applicable in economic evaluations in healthcare sectors that fall outside the remit of pharmaceutical care. For these reasons the National Health Care Institute (Zorginstituut Nederland, ZIN), successor to CVZ, decided in 2014 to compile one revised and updated version. A non-permanent committee was established to support the National Health Care Institute in compiling the new guideline.

The key objective was to develop a new guideline unifying the Dutch guidelines available at the time, linking up with international developments, providing unambiguous instruction and accommodate its use in health fields beyond that of pharmaceutical care. Examples include evaluations performed to inform decisions on whether new healthcare interventions should be implemented or reimbursed or, on the other hand, existing interventions should be discontinued or no longer be reimbursed.

The new version is made up of different sections: the main body of the document containing the actual text of the guideline, and a number of in-depth modules. Currently these in-depth modules address the topics 'Uncertainty and Value of Information', 'Quality of Life', 'Budget Impact Analysis', and one is a revision of the previous cost research manual from 2010. These modules elaborate on specific elements of the guideline. It is foreseen that in-depth modules on other topics will be added from time to time. The new guideline specifies – on the basis of a reference case – the minimal requirements which all economic evaluations should meet. A separate chapter describes its broader application in health fields such as prevention, diagnostics, medical devices, Long-term care and forensics.
Until the 1st of July 2016 the National Health Care Institute will still perform economic evaluations in the context of a drug reimbursement file on the basis of the Guidelines for pharmacoeconomic research, updated version from 2006. As from that date the newly updated guideline will be in force.

Introduction

History

In 1999 the Health Care Insurance Board (College voor Zorgverzekeringen, CVZ) issued the 'Guidelines for pharmacoeconomic research'. An updated version (2005) was entered into force in April 2006. This document provided an analytic framework for the pharmacoeconomic evaluations which are part of a reimbursement file. The framework serves to assess whether there is sufficient evidence for the efficacy of a drug for which reimbursement has been requested. Previously, three sets of guidelines were available, namely the above-mentioned Guidelines for pharmacoeconomic research (2006), the Guidance on outcome research (2008) and the Cost research manual (2010). It should be noted that these guidelines placed a primary focus on pharmaceutical care. Nevertheless, the Guidelines for pharmacoeconomic research and the Cost research manual were also used for economic evaluations of other types of care in the Netherlands.

Aim of the new Guideline

This new Guideline unifies all previously available Dutch guidelines, links up with international developments, provides unambiguous instructions and is broadly applicable, not only to pharmaceutical care. The Guideline is intended for those who perform economic evaluations to inform decisions on whether new healthcare interventions should be implemented or existing interventions that should be discontinued.

Non-permanent committee

A non-permanent committee was established, charged with advising the National Health Care Institute on the revision of the guidelines. The committee was composed of: Prof. M.J. IJzerman (chairman), Dr. M.J. Al, prof. A. de Boer, Prof. W.B.F. Brouwer, Prof. J.J. van Busschbach, Dr. M.G.W. Dijkgraaf, Dr. T.L. Feenstra, Dr. M.A. Joore, Prof. J. Kievit, Prof. J.J. Polder, Dr. G.A. de Wit and Dr. M.T. Groot (observer).
The following staff members of National Health Care Institute carried the project: Dr. S. Knies (project leader/secretary), N.M. Dragt MSc, J.J. Enzing MSc, Dr. L.A.H. Hermsen and Dr. S.M.C. Vijgen.
Two observers from the ministry of Health, Welfare and Sport were involved: Dr. J. Prins and A.A. Golja MSc.

Document structure

Guideline

This document is made up of different sections: the main body of the document, which contains the actual text of the Guideline and a number of in-depth modules. Currently in-depth modules are available addressing the topics 'Uncertainty and Value of Information', 'Quality of Life', 'Budget Impact Analysis' and Costing Guidance. These modules provide more detailed information or further elaborate on the topic in question. Note that an in-depth module is an integral part of the Guideline and has the same status. Moreover, the way in which the document is structured permits adding in-depth modules on other topics without having to re-organize the entire document.

Main body document

The main body of the document presents six content-based chapters and an overview of the requirements of the reference case which all economic evaluations must meet. Chapter 1 deals with the design of an economic evaluation and chapter

2 places a focus on the methodology. Chapter 3 addresses the different input data of an economic evaluation. The reporting of the results of an evaluation is elaborated in chapter 4. Chapter 5 exclusively deals with budget impact analyses. Lastly, chapter 6 pays attention to aspects to be considered in the context of other fields of application than pharmaceutical care.

Amendments to the current guidelines
In comparison with the current guidelines the new Guideline includes many amendments. The major content changes are listed in table 1.

Table 1: Overview of amendments in the new Guideline in comparison with current guidelines

	Previous guidelines (2006)	New Guideline (2016)
Topic		
Time horizon	Chosen such that a valid and reliable conclusion can be drawn	Preferably a lifelong time horizon
PICOT scoping	-	Scoping beforehand with relevant stakeholders
Type of economic evaluation	CUA/CEA/CMA	Always CUA
Costs	Costs: subdivided as direct versus indirect and medical versus non-medical	Costs: subdivided as described in Drummond et al. 2015
	Indirect medical costs not taken into account	Medical costs in the life years gained (indirect medical costs) taken into account
QoL and QALY	Generic measure, preferably EQ-5D	EQ-5D-5L with Dutch valuation
Budget impact analysis	-	Described in chapter 5
Modules		
Costing guidance	-	Chapter presenting reference prices of the most common healthcare cost units
	-	Excel file with overview of reference prices of the most common healthcare cost units: available online

Background literature
The Guideline is not intended to be a handbook or a reference book for the conduct of economic evaluation. It is assumed that the user of the Guideline has some basic knowledge of economic evaluations. Please find below an overview of handbooks that may be useful.

- Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
- Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes. (fourth edition) Oxford University Press. 2015
- Drummond MF, McGuire A. Economic Evaluation in Health Care; merging theory with practice. Oxford University Press, 2001.
- Glick H, Doshi J, Sonnad S, Polsky D. Economic evaluation in clinical trials. Oxford: Oxford University Press; 2007

- Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in health and medicine. 1996. Oxford University Press.
- Hunink MGM, Weinstein MC, Wittenberg E et al. Decision making in health and medicine. Integrating evidence and values. Cambridge Medicine, 2014.
- Rutten-van Mölken M, Uyl-de Groot C, Rutten F. Van kosten tot effects. Een handleiding voor evaluatiestudies in de gezondheidszorg. Elsevier gezondheidszorg, Tweede herziene druk, 2010.

Reference case

As the purpose of this Guideline is to improve the comparability of economic evaluations conducted in the Netherlands, it starts with an explanation of the reference case which all economic evaluations at minimum should comply with. This refers in particular to the methods which are of importance in the view of the National Health Care Institute. It may be necessary to deviate from the reference case, but the reasons and nature of deviations should be well substantiated and clearly described. It should be noted that an alternative analysis cannot replace the standard analysis. After all, not conducting the standard analysis carries the risk of more methodological uncertainty, which ultimately might affect the comparability and assessment of studies. Some deviations from the standard analysis may be inevitable for an economic evaluation of an intervention in a field other than pharmaceutical care or cure/curative care, such as prevention, diagnostics or long-term care. Chapter 6 addresses the use of the Guideline in the evaluation of healthcare interventions in these other fields in more detail.

Part	Reference case	See Guideline/ paragraph
Perspective	Societal perspective	1.2
PICOT	P: (Patient) population in the Netherlands	1.3
	C: Standard of care or usual care in the Netherlands	
	T: Lifetime	
	Scoping beforehand with relevant stakeholders	1.4
Economic evaluation type	Cost-utility analysis (CUA)	2.1.1
Discounting	Costs 4% and effects 1.5%	2.2
Uncertainty and sensitivity analysis	Univariate, probabilistic sensitivity analysis and scenario analysis	2.3
Effectiveness	Systematic review	3.1
Costs	All costs inside the healthcare sector, patient and family and other sectors. Productivity losses: friction cost method. For reference prices use wherever possible the 'costing guidance' module	3.2
Effects	Expressed in QALYs, a least the EQ-5D-5L with Dutch valuation and whenever relevant also life years gained	3.3
Results reporting	Total costs and effects, incremental costs and effects, incremental cost-effectiveness ratio	4.2
	Univariate sensitivity analysis: Tornado diagram and table Scenario analysis: table Probabilistic sensitivity analysis: CE-plane and CEAC	4.3

1 Chapter 1 Framework of the economic evaluation

Before an economic evaluation of a healthcare intervention can be performed, the researcher or the person who will perform the evaluation needs to provide an evaluation framework and design. It is important to note that all aspects must be clearly described in the reporting and that all choices are justified.

1.1 Objective and user

The first step in an economic evaluation is to clearly state what decision-making problem the evaluation will solve. Or in other words: defining the **objective.** The objective should be linked to the decision-making problem of the future user(s) of the evaluation. This is why it is important to also define the **user(s**) of the evaluation. Apart from the National Health Care Institute, possible users include health insurers and health facilities.

1.2 Perspective

When conducting an economic evaluation it is imperative to follow the reference case. This means that the economic evaluation is carried out and reported from the societal perspective. All relevant societal costs and benefits, irrespective of who bears the costs or to who the benefits go, should therefore be taken into account in the evaluation and reporting.

In addition to the societal perspective the results can be presented from other perspectives (such as the healthcare perspective). The relevance of such a choice should be clearly underpinned.

1.3 PICOT or research question

Once the objective, the user and the perspective of the evaluation have been clearly defined, the next step is to formulate the research question for this specific economic evaluation. The evaluation is conducted on the basis of the research question formulated in accordance with the PICOT criteria, as described by the Centre for Evidence Based Medicine[1]:
- Patient = the patient or target population;
- Intervention = the intervention to be assessed;
- Control = the comparative intervention(s);
- Outcome = the relevant outcomes/outcome measures;
- Time = the relevant time span for which effects and costs must be measured

Table 2 provides an overview of all aspects that should be considered in the framing of the PICOT. It is important to note that all parts must be clearly described in the reporting and that all choices are justified.

Table 2: Aspects to be considered in the framing of the PICOT

What intervention in which patient population?
What is the intervention under scrutiny compared with?
– standard and/or usual care
– direct or indirect comparison
What are relevant outcomes?
– justify choice of outcomes
– favorable and unfavorable outcomes
– choice of relevant measurement instruments
What outcome of the comparison is satisfactory?

> – more effective and equally effective
> – clinically relevant difference
> What is the minimum required time horizon for the analysis?

1.3.1 Patient population (P)

The patient population or target population is the group for which the intervention is intended in the Dutch clinical practice. Where available, Dutch epidemiological data should be used to describe the patient population as well as the relevant subgroups. Differences in age, gender and disease characteristics between and within populations can vary and seriously influence the results. In the event that the cost-effectiveness with respect to effects and/or costs differs between subgroups, separate subgroup analyses are an option. Note that a-priori hypotheses should have been generated for this situation if the economic evaluation is part of a randomized controlled trial (RCT).

1.3.2 Intervention (I)

The intervention under consideration is an intervention that in the Dutch practice can be applied in the defined patient population. For both the intervention and the comparative treatment it is essential to take into account how interventions in the Dutch practice are, or will be, implemented. Important aspects in this respect include the setting in which the interventions will be carried out and the (expected) uptake by healthcare providers and patients.

1.3.3 Control or Comparator (C)

As described in the reference case the intervention should be compared with the standard of care and/or usual care.i The standard treatment is the intervention which in daily practice or in accordance with clinical guidelines is considered the treatment of first choice. Usual care includes care procedures which in clinical practice are routinely applied. If the standard treatment is not part of the usual care or cannot be defined, usual care procedures may be included in the analysis, whether or not next to the standard treatment.

With respect to all interventions, the most recent national and international guidelines and standards should underlie the choice of the comparative intervention. In some cases multiple interventions which are well comparable to each other in terms of outcomes might qualify for a specific patient population. The concept of 'usual care' may differ from country to country or even within one country. Furthermore, usual care may also imply that no treatment is offered, but that instead a wait-and-see policy is recommended, whether or not in combination with symptom relief (supportive or palliative care).

Relevant sources:
Detailed information about the selection of the comparative intervention can be found in the revised report 'Stand van de wetenschap en praktijk' ('Established medical science and medical practice') from 2015.[2] In addition, a variety of international documentation is available from EUnetHTA: "Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures". Accessible via www.eunethta.eu/eunethta-guidelines.[3]

1.3.4 Outcomes (O)

In effectiveness research the value of the intervention under scrutiny as compared to the usual care is determined using pre-defined outcome measures. Different

types of outcome measures can be distinguished. Clinical outcome measures reflect morbidity, mortality or quality of life. Patient-reported outcome measures (PROMs) refer to outcomes directly reported by patients themselves, such as pain scores or quality of life. Occasionally it may be necessary to use intermediate or proxy outcome measures, because it is not possible to set up a study with sufficient robustness to demonstrate differences in hard outcome measures (for example mortality). A prerequisite for intermediate or surrogate outcome measures is a proven sensitivity to change in the clinical outcome. Another consideration is the analytic accuracy of the test with which the intermediate or surrogate outcome is measured. Composite outcome measures can be used when two outcomes are considered that are clearly clinically related. These are often used for outcomes whose frequency is low (for example mortality). Another composite measure is the QALY, which combines life years with quality of life.

The choice of a specific outcome is determined by factors such as the eligible patient population, the health condition and the aim of the treatment. It is recommended to use validated outcome measures[1] which are clearly related with the condition and the intended purpose of the treatment. Outcome measures should bear relevance to the healthcare intervention and therefore it is important to select these before the start of the evaluation study in consultation with the professional groups of caregivers involved and with patient organizations.

As indicated in the reference case, the QALY is the standard outcome measure in healthcare economic evaluations. The outcomes of an economic evaluation consist of clinical outcome measures and measures that describe and value the quality of life on one side, and the costs of treatment on the other. These are the outcomes that form the basis of the economic evaluation.
More information on the recommended outcome measures for quality of life is provided in section 3.3 and the 'Quality of Life' in-depth module. More information about the cost (categories) to be use and the quantification of healthcare consumption is given in section 3.2 and the cost guidance module.

1.3.5 *Time horizon (T)*
The time horizon required for an economic evaluation should preferably cover the expected lifetime. A different time horizon may be chosen, however, if well-argued reasons can be provided that a lifetime horizon is not relevant. In that case the time horizon should allow for a valid and reliable pronouncement on the differences between the effectiveness and costs, respectively, of the interventions that are being compared.

Additional information on the design of a framework for an economic evaluation can be found in a variety of literature.[4,5]

1.4 **Scoping of the PICOT**
It is recommended to 'scope' the PICOT beforehand with the relevant professional scientific organizations and patient organizations to benefit from their expertise and experience. If an economic evaluation is conducted at the request of or for the purpose of an assessment by the National Health Care Institute, the National Health Care Institute must be consulted before the PICOT is drawn up. Detailed information on scoping is available in the revised report 'Stand van de wetenschap en praktijk' from 2015.[2]

---

[i] Internal validity, reproducibility and sensitivity of an outcome measure is based on scientific publications.

2 Chapter 2 Method: analytic approach

2.1 Analytic techniques

2.1.1 Types of economic evaluations

A number of different analytic techniques in which two or more treatments are compared with one another are available for use in an economic evaluation. The most frequently used techniques are the cost-utility analysis (CUA) and the cost-effectiveness analysis (CEA). A budget impact analysis (BIA) might be added to the economic evaluation. A BIA provides information about the financial consequences of the intervention. Specific guidelines are in place for the design of a BIA (see chapter 5).

Both in a CEA and in a CUA the differential costs (incremental costs) are set off against the differential effects (incremental effects) of the treatments that are being compared. A CEA makes use of (clinical) effect sizes (for example millimeters mercury to express blood pressure, event free survival, or life years gained). In a CUA the effects are expressed in a generic measure of the quantity of life years corrected for quality of life, or Quality Adjusted Life Years (QALYs). These QALYs are determined with the aid of generic measurement instruments such as the EQ-5D (see section 3.3 and the in-depth module about quality of life). A CUA allows for better comparability of the outcomes of analyses between different health conditions. Conducting a CUA is therefore standard when performing an economic evaluation in the healthcare sector (see reference case).

2.1.2 Analytic approach

The costs and effects of interventions can be calculated using a decision model. This approach constitutes the benefit of making comparisons in a complex situation in a transparent and consistent manner. In the case that no single comparative study is available that fulfills the PICOT, see chapter 1, the only alternative is a decision model.

The choice for either an empiric approach (in which all relevant costs and effects are collected in one clinical study) or a model-based approach should be well justified. In a model-based approach, the choice of a specific technique much depends among other things on the research question and the medical condition. The most commonly used models are: decision trees, 'state-transition' models (also known as Markov models), 'discrete event' simulations, and 'dynamic transmission' models. Recommendations for the choice of the most appropriate technique can be found in the ISPOR/SMDM guidelines on modelling.[6]

2.2 Discounting

If effects and costs data are collected or presented for a time horizon exceeding one year, the effects and costs generated after the first year need to be discounted. In the Netherlands, the costs are discounted at a different rate than are the effects: the costs at a constant discount rate of 4%; the future effects at a constant discount rate of 1.5%.

2.3 Uncertainty analyses

The results of an economic evaluation, obtained through either a model-based or an empirical approach, are surrounded by uncertainty. Uncertainty analyses have a

two-fold aim:
1) to determine the degree of uncertainty around the mean costs, effects and cost-effectiveness ratio;
2) to quantify the consequences of the uncertainty, and to determine if and how additional research aiming to reduce the uncertainty can be of value.
The first aim is reached by performing uncertainty analyses; the second by conducting a value of information (VOI) analysis.

Uncertainty can result from differences in opinion on the methods to be used, the methodological uncertainty. This type of uncertainty is overcome by following this Guideline and the reference case. Other sources of uncertainty are: variability in outcomes between otherwise identical patients (can be partly explained; heterogeneity), estimates of data on the basis of limited information from (sample) data, structure of the decision model or regression model, and missing data. As the methods to deal with the different sources of uncertainty differ between an empirical analysis and a model-based analysis, these are described separately below.

Relevant sources:
Detailed background information about variability, uncertainty and heterogeneity can be found in, for example, the following sources: ISPOR/SMDM Taskforce on modelling, report 6[7]; Drummond et al.;[8] Bilcke et al.;[9] Briggs et al.;[10] Ramsey et al.[5]

2.3.1   *Uncertainty analyses in empirical economic evaluation*
Insight in the consequences of variability can be obtained by demonstrating the presence or absence of heterogeneity. In an empirical approach, subgroup analyses should be performed for all pre-defined subgroups.

In the design stage of a study, the probability of uncertainty as a result of limited information might be reduced by including a sufficiently large sample and by, for example, limiting the possibility of missing values.

If a study has already been performed, the uncertainty around the total and incremental costs, effects and the incremental cost-effectiveness ratio should be quantified with statistical methods. The method used most frequently used to determine the confidence intervals for total and incremental costs and effects and the incremental cost-effectiveness ratio (ICER) is 'bootstrapping'.[11] The number of sampling draws used in the bootstrapping should demonstrably lead to a stable result. Confidence intervals around the ICER often do not provide the desired information, for example if outcomes are distributed over multiple quadrants of the cost-effectiveness-plane (CE-plane).

Importantly, the statistical analyses used to evaluate the data must take into account any missing values, to prevent that outcomes such as means and confidence intervals are incorrect or biased.

The literature contains information on both naïve (e.g. complete cases, mean imputation and last value carried forward) and statistically appropriate methods (e.g. multiple imputation and EM algorithm) to correct for missing values. In economic evaluations 'state of the art' methods are needed to correctly assess the uncertainty generated by missing data, and the multiple imputation method is often the optimal method. Subsequently, the question whether the application of a different suitable method would lead to large differences in both the point estimations and the confidence intervals should be explored through scenario analyses.
Regression techniques may serve to increase the precision and to correct for

differences between subgroups. If the mean costs and effects are determined by regression techniques, the regression technique chosen and the definition of the regression model should be well substantiated. The possible impact of assumptions in the regression analysis on the outcomes should be demonstrated in sensitivity analyses, and likewise the impact of varying the regression parameters within their 95% confidence intervals. Uncertainty around fixed parameter values, such as unit prices, should be demonstrated in sensitivity analyses.

Relevant sources:
Further information is found in Ramsey et al. (2015).[5] Methods to correct for missing data are explained in Sterne et al. (2009)[12] with emphasis on imputation techniques and in Oostenbrink and Al (2005)[13] more specifically for cost data.

2.3.2 Uncertainty analyses in a model-based economic evaluation

Variability will come into play when a patient-level model is used. Importantly, the simulated population size should demonstrably lead to a stable result. Furthermore it is important to clearly discriminate between the variability and the parameter uncertainty. In the event that heterogeneity is of relevance, subgroup analyses should be performed and described. Any heterogeneity should be detected in subgroup analyses, and 'scoping' with the stakeholders is leading in the choice of the subgroups.

The influence of parameter uncertainty should be examined using probabilistic sensitivity analyses (PSA) in which all uncertain parameters are analyzed. The distributions used must be reported and justified. The choice of the distribution depends in particular on the logical limitations of an input parameter.[10] The degree of variance in the distribution need be justified as well, based where possible on observed variance in studies (e.g. standard errors and confidence intervals). The number of iterations used in the PSA should demonstrably lead to a stable result. The ICER should be calculated by dividing the mean costs by the mean effects across all PSA iterations.
Furthermore, univariate and multivariate sensitivity analyses should be performed to provide insight in the relative influences of input parameters on the ICER and in the consequences of fixed values in the model such as discount rates and prices.

It is imperative to bring to light the implications of structural assumptions. The structural uncertainty can be made transparent by presenting a clear overview of the model assumptions. In addition the choices made should be underpinned. The influence of the assumptions on the results of a decision model should be explored with the aid of scenario analyses.

2.3.3 Value of information analysis

A value of information (VOI) analysis is based on the results of the probabilistic sensitivity analysis. The expected value of perfect information (EVPI) initially quantifies all consequences of the uncertainty around the parameters. In all model-based economic evaluations where decision uncertainty is found (when the probability that the intervention is cost-effective is lower than 100% at the applicable reference value but still higher than 0%) an EVPI should be calculated.

Supplementary VOI analyses can answer the question whether, and what kind of additional study is efficient. Additionally, they can indicate whether the definitive decision should be postponed while awaiting the results of additional studies. These analyses are relevant if additional study and/or postponement of the decision is a viable policy option. The in-depth module 'Uncertainty and VOI' elaborates on the significance and the application of VOI analyses. This module also provides a

flowchart specifying what type of VOI analyses should be performed and when.

2.4 Missing data

Data can be missing due to various reasons. As already pointed out in paragraph 2.3, missing data are integral to the statistical uncertainty in an empirical analysis. This section addresses the implications of missing data in decision models.

2.4.1 *Missing data in model-based economic evaluations*

In an economic evaluation based on a decision model, missing data in an empirical study have repercussions for the estimation of parameter values. Another thing that might happen is lack of evidence for an estimation of a parameter value. This is never a reason, however, to leave this parameter out in the modelling. In this instance assumptions should be made in the model as recommended by a panel of experts.

2.4.2 *Extrapolation techniques*

One type of missing data which is quite relevant in the context of model-based evaluations concerns the long-term projection. If data for a longer time horizon are missing, analyses can only be performed using extrapolation techniques, thus allowing for the determination of the total survival gain. Longitudinal time-to-event data (for example survival data or time-to-relapse data) are primarily extrapolated by assuming a parametric distribution. Then, extrapolation is achievable by estimating the parameters of this distribution on the basis of the observed data. The uncertainty around these parameters should as well be included in the PSA. Commonly used parametric methods or models are exponential, Weibull, Gompertz, log-logistic and lognormal, but also more complex and more flexible methods are available.

As the different methods may lead to different outcomes, the choice for a certain method should always be well substantiated. Furthermore, possible different outcomes when other distributions are used should be explored in scenario analyses. Importantly, the reporting should include the results of the statistical tests used to compare the goodness of fit of the different distributions with the observed data.

Relevant source:
Survival analysis methods are explained in more detail in NICE DSU technical support document 14: Survival analysis for economic evaluations alongside clinical trials- extrapolation with patient-level data.[14]

2.5 Evaluation of validity

Generalisability of results

Both empirical studies and evaluations on the basis of decision models should give insight in the relevance of the outcomes for the specific situation in the Netherlands. More specifically, this concern both the generalizability of specific results of empirical studies to the general population and the validity of the decision model used.

Validation of decision models

The decision model used in a model-based economic evaluation must have been validated to ensure that the results are usable, reliable and credible. The validation should encompass not only the conceptual model, but also the input data, the software-aided implementation (code) and the model outcomes.
The reporting should inform on the validation method and the obtained result. Optimal reporting on the validation will be achieved not only by using a purpose-

designed assessment tool as a checklist for the relevant steps and tests but also by providing a systematic representation of the validation steps. An example of such a tool is AdVISHE.[15]

3 Chapter 3 Input data

3.1 Effectiveness

An economic evaluation may be performed in the context of an empirical study (piggy-back study alongside an RCT or observational study) or a model-based study. If the context is an empirical study, the input data on effectiveness are derived from the study. Supplementary effectiveness data, if necessary, should come from other relevant publications.

A model-based economic evaluation should be grounded on data of all relevant studies with a satisfactory scientific quality. The research question determines how these studies should be identified and what level of evidence of effectiveness is minimally required. For this reason the research question must be defined as precisely as possible, see section 1.3: PICOT or research question.

As recommended in the recent report '*Stand van de wetenschap en praktijk'*, prior to collecting relevant literature or designing an effectiveness study, one should explore the desired and feasible characteristics of the study in light of the intervention and indication.[2]

3.1.1 Literature search

A systematic search of relevant publications should adhere to a pre-defined protocol. The search strategy as well as the databases used and any publication date restrictions must be reported (in an appendix) for the purpose of verification. An overview must be provided of all selected studies and the selection criteria used. Criteria for inclusion or exclusion of studies should be reported.

3.1.2 Relevant studies

On the basis of the results of the literature search an overview of existing evidence should be provided as well as a quantitative effect estimate. The latter may be based on published studies and/or new empirical data. In all cases the method employed should be systematic and complete and aimed at preventing bias as much as possible.

Systematic review

In a model-based economic evaluation the clinical effectiveness data need to be underpinned by a systematic review of the literature, provided this is of sufficient quality. If this is not the case, it is recommended to perform a systematic review and to systematically evaluate the collected scientific evidence with the aid of the GRADE approach, or: 'Grading of Recommendations Assessment, Development and Evaluation.' The GRADE-method is a systematic, integral and transparent method to evaluate and grade the quality of scientific evidence. For detailed information about the GRADE-method see the report '*Stand van de wetenschap en praktijk*'[2], and the GRADE handbook.[16]

Randomized studies

It is strongly recommended to base the systematic review on randomized studies of sufficient quality.[17] A well designed and well performed (double-blind) randomized controlled clinical trial (RCT) has in theory the least probability of confounding of the effect and consequently carries the highest degree of certainty about the causal relation between the intervention and the observed effect (irrespective of this effect being favorable or unfavorable). Other study designs are less reliable in that they have a greater probability of bias (confounding, information- and selection bias), i.e.

the possibility that involvement of other (unknown) factors may lead to under- or overestimation of the difference in effect between the intervention and the comparative treatment. The GRADE handbook16 extensively describes how to evaluate the methodological quality or risk of bias of randomized studies and how to perform the GRADE-evaluation.

Non-randomized studies
Results of RCTs are not always available in the literature. In addition, an RCT also has it downsides and limitations. Certain situations may arise, therefore, in which non-randomized or non-comparative studies will suffice or even may be preferred, for example when there is a clear dose-response relation, the natural course of a condition is known or in the case of rare diseases. A supplementary observational study may suffice to identify long-term effects or side-effects and may be preferred if the purpose is to measure (results of) healthcare consumption in daily practice, rather than under strictly controlled research conditions.[18,19]
When non-randomized studies are utilized, such as observational studies, potential bias must be fully explored and reported and be taken into account in the economic evaluation. The GRADE handbook16 describes how to evaluate the methodological quality or risk of bias of non-randomized studies and how to perform the GRADE-evaluation. Apart from the GRADE handbook[16], two other commonly used instruments to assess the methodological quality or risk of bias of non-randomized studies are the Newcastle Ottawa Scale[20] and the Down and Black Instrument.[21]

3.1.3   *Direct versus indirect comparisons*
It is strongly recommended to perform the evidence synthesis on the basis of randomized studies using direct (head-to-head) comparisons, as this approach produces the most reliable evidence for the value of a certain intervention in comparison with the standard care. Even if only a single randomized study of sufficient quality and with a direct comparison is available, this should then be used.

However, this ideal situation is not commonplace. In such cases, indirect comparative studies may be used. In an indirect comparison, results of study arms from different studies are compared with each other. These are preferably study arms from randomized studies with indirect comparisons and from non-randomized studies, such as observational cohorts, whereby one study arm with a relevant patient group is selected.

When executing indirect comparisons one should select studies carefully to prevent bias as much as possible. It is recommended to use 'state of the art' methods for integrating indirect comparisons in a single quantitative analysis: either Bayesian[22] or frequentist[23], that is, a mixed treatment comparison (MTC) or a network meta-analysis. Please note that here too the quality of the evidence is preferably evaluated with the GRADE-approach.[24]

Placebo-controlled studies
Placebo-controlled studies are in principle not suited for economic evaluations, because in clinical practice placebo is considered neither standard care nor care as usual. Nevertheless, placebo-controlled studies may be included in a mixed treatment comparison or in a network meta-analysis to complete the network.
If placebo is taken to be a proxy for 'best supportive care', which indeed may be usual care, this approach should be well justified.

3.1.4   *Sources for clinical effectiveness*
Clinical effectiveness data need to be underpinned with a systematic review of the literature. The clinical evidence should be based on the most appropriate source,

given the research question. Studies with the highest possible level of evidence and the least (possible) risk of bias are preferred. The quality of all studies should be assessed and reported.

For detailed information about, the systematic evaluation of the effectiveness of an intervention, the GRADE approach and the inclusion of indirect comparisons, among other things, we refer to the report '*Stand van de wetenschap en praktijk'*, updated in 2015.[2]

Relevant sources:
Additional information about mixed treatment comparison in found in the Cochrane handbook[25] and the ISPOR taskforce report.[26] The Cochrane site offers a complete overview of tools and software for the conduct of network meta-analyses: http://cmimg.cochrane.org/network-meta-analysis-toolkit.

Two often used software programs are:
Bayesian: http://drugis.org/software/r-packages/gemtc
Frequentist: http://cran.r-project.org/web/packages/netmeta/index.html

3.2 Costs

Please consult the 'Manual for cost research' ((Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economic evaluations in de gezondheidszorg (2015))[27] for guidance on the identification, the measurement and the valuation of the costs. The use of this manual promotes uniformity and standardization of the cost identification, measurement and valuation in economic evaluations of healthcare interventions.

3.2.1 *Cost identification*

Economic evaluations should be conducted from the societal perspective. This means that all relevant costs and effects must be included, irrespective of who meets the costs. This does not just refer the costs inside the healthcare system but also to the costs for the patient and the family (for example incurred by informal care), as well as the costs incurred in other sectors of society (such as education or the judiciary) and costs as a result of productivity losses.

All categories of costs need to be indicated and explained separately. Specific attention should be paid to avoid double counting.

If in addition an evaluation is carried out from another perspective than the societal perspective, all relevant costs and categories of costs for that perspective should be clearly defined and included. It may also be necessary to adjust the valuation to the other perspective.

Table 3 provides an overview of the categories of costs that are relevant to the societal perspective as well as the categories of costs related to the healthcare perspective.

Table 3: Perspectives and associated cost categories

PERSPECTIVE	COST CATEGORIES	SPECIFICATION
Societal perspective	Costs within the healthcare system	- All medical costs directly resulting from the intervention - All medical costs incurred during life years gained
	Costs for patient and family	For example: - Travel expenses

		- Own contributions/co-payments
		- Time spending costs
		- Costs of informal care
	Costs in other sectors	Costs incurred in sectors outside the healthcare system, for example municipal services, education or voluntary work
		Productivity losses: costs of absenteeism or unproductivity during paid work (presenteeism) and unpaid work
Healthcare perspective	Healthcare sector	- All medical costs directly resulting from the intervention
		- All medical costs incurred during life years gained

Bron: Drummond et al. (2015)[8]

Costs are quantified using the formula: costs = volume*price. Volume is obtained by measuring resource use or units. To calculate the costs these units or the resource use is then multiplied with the price per unit.

3.2.2 Volume measurement

After the identification of the categories of costs to be included in the analysis, the next step is measuring the volumes (for example healthcare consumption). This measurement comprises an inventory of the deployment of people and resources during, after and sometimes also before the intervention. Volume data derived from international studies must be validated for the Netherlands.

3.2.3 Cost valuation

Subsequently, the identified and measured volume units need to be assessed in monetary units. Next to guidelines for carrying out cost-price research, the 'cost manual' gives concrete reference prices for a number of common units. [27]

3.2.4 Categories of costs

For evaluations taking a societal perspective, three categories of costs can be distinguished, namely costs within the healthcare system, costs for patient and family and costs in other sectors.

Costs within the healthcare system

The category costs within the healthcare sector comprises two types of costs, which both must be included in the economic evaluation:

- All medical costs directly related to or influenced by the prevention, diagnostics, therapy, rehabilitation and management of the disease or treatment. The costs of the treatment of side-effects of an intervention are also part of this category.
- All medical costs which may arise during gained life years, also referred to as indirect medical costs. This pertains to all medical costs arising as a consequence of the life-lengthening effect of an intervention. If an intervention increases the life expectancy, this will normally lead to additional medical costs, either 'related' costs (e.g. anticoagulants after cardiac surgery) or 'non-related' costs (e.g. the costs of hip fracture surgery in the life years gained after a heart transplant). Both types of costs should be taken into account in an economic evaluation of a life-lengthening intervention. More information about calculating costs in gained life years (also non-related costs) can be found in the 'costs manual' module.

Costs for patient and family
The costs for a patient and family may include, among other things, travel expenses, own contributions/co-payments or adaptations to one's home. One of the major cost items in this context concerns the time spent on the intervention. For example the time needed to undergo treatment or the time involved in the provision of informal care. The time spent on informal care may be at the expense of paid work, unpaid work and leisure time. More information on the valuation of informal care and time spending costs of patients can be found in the 'costs manual' module.

Costs within other sectors
The costs incurred in other sectors are strongly dependent on the intervention to be evaluated. Examples include costs made by a local council for welfare work, costs of special education or the judicial system, or the costs of volunteers. It is important to be complete and yet avoid double counting. More information about the valuation of costs within other sectors is available in the 'costs manual' module. More information can also be found in the guide on intersectoral costs and benefits: www.maastrichtuniversity.nl/web/Main/Sitewide/Content/ClassificatieIdentificatieEn WaarderingVanIntersectoraleKostenEnBatenVanPreventieveInterventies.htm.[28]

If illness or treatment prevents people from being productive, this will incur costs. Not only paid productivity (paid employment) is important in this respect, but also unpaid work. In both cases the productivity losses (or -gains) involved must be specified and valued. In the case of paid employment a distinction can be made between less productivity during the job ('presenteeism') and work absenteeism, which both may be of importance.
The friction cost method is recommended to calculate productivity losses as a result of paid work absenteeism. This method is based on the idea that sick employees who are absent for a long time can be replaced. Productivity losses will then mainly occur during the time an employer needs to replace a sick employee, the friction period. A friction period of 85 calendar days (12 weeks) is applied in the Netherlands. More information and explanation on the calculation of productivity losses using the friction cost method is available in the 'costs manual' module'. An other method is the 'Human Capital' method, but this is not recommended.

3.2.5   *Adjustment for inflation/price indexation*
In each economic evaluation it is essential to mention the year to which prices refer. In the event that cost prices of different units or volumes are not from the same calendar year, the prices should be corrected for inflation between the years. This is done with the use of the consumer price index (CPI), available in the CBS StatLine database (Centraal Bureau voor de Statistiek – Statistics Netherlands). More information and explanation is available in 'costs manual' module.[27]

3.2.6   *Sources for cost data*
Both the volumes and the cost-prices per unit should be based on national data wherever possible. The 'costs manual' module extensively addresses methods for the measurement of volumes and standard cost-prices.[27] If in the absence of national cost data, data from international publications are used, these must be validated for the Dutch setting. One way to accomplish this is to request input from an expert panel.

3.3   Quality of life

3.3.1   *Quality Adjusted Life Years (QALYs)*
Interventions may have an effect on the life expectancy, quality of life, or both. Both types of effect are essential in economic evaluations and both can be used to justify

the costs. The relevance of the effects on life expectancy and quality of life relative to the costs should be demonstrated by means of a 'cost-utility analysis' (CUA). The 'utilities' of the CUA are operationalized by a parameter reflecting the amount of life years gained in combination with quality of life, or 'Quality Adjusted Life Years': QALYs.

3.3.2 Correction factors of quality of life
In the context of measuring QALYs, quality of life is determined by means of validated, generic quality-of-life questionnaires completed by patients. The outcome measures of these questionnaires are 'health states' defined in terms of health-related quality of life. The validated questionnaires assign a 'quality-of-life correction' to the health states. This correction factor is referred to as a 'weighting', 'valuation' or 'utility'. Basically this implies the 'quality of life' is offset against 'expected life expectancy'. The patients' health states should be valued based on the preferences of the general population, so that the analyses can be performed from the societal perspective.

3.3.3 Reference case
In view of the possible differences in the assessment of quality of life that may arise from the use of different questionnaires, quality of life should consistently be measured with the EQ-5D-5L and should be valued using Dutch reference values.[29] This may be taken as a simplification of reality, however, and consequently may not always be appropriate. Alternative questionnaires and other methods to evaluate the quality of life may therefore be added next to the reference case. These alternatives do not replace the reference case but may be required to identify the effect to its fullest. This means that even when the EQ-5D-5L is expected to be not sensitive, the EQ-5D-5L should still be administered. The alternatives are addressed in the in-depth module on quality of life.

3.3.4 Reporting
In the event that alternatives are presented next to the reference case, these alternatives need to be justified and their validation should be described. In a decision model using health states the valuations of each health state should be reported separately. If QALYs are estimated on the basis of an 'area under the curve' analysis, the course of the mean quality of life over time should be presented in a table or graph. If the valuations of quality of life are derived from the literature, the following characteristics should be reported separately for each valuation: the questionnaire or valuation method; nationality of the respondents and the perspective: patient valuations or societal valuations.

3.4 Expert opinion
Where data are clearly lacking, it is recommended to ask a panel of experts to provide data for 'input' in a model or for the design of a model. One of the common methods to organize expert panels is the Delphi method. If an expert panel is consulted, the way in which the experts were selected must be clearly explained so that their expertise and independence can be verified. The following information should be provided: the data collection method, the process through which consensus, where needed, was achieved, and the analysis method. Furthermore, the data provided by the expert panel should be specified and the impact of uncertainty on the results must be elucidated through sensitivity analyses.

4 Chapter 4 Reporting

4.1 Reporting of input values

4.1.1 Reporting of effectiveness data
If the effectiveness input data are based on specific clinical studies, details of these studies should be reported such as patient characteristics (among others age, gender, baseline values of clinically relevant parameters) and the effectiveness of the intervention. Effectiveness may be reported based on the pre-defined clinically relevant outcome measures and minimal clinically relevant difference or maximal clinically acceptable difference.

If estimations are based on systematic reviews and meta-analyses, the methods used should be reported in a transparent manner.

4.1.2 Reporting of costs
The prices and volumes of all cost components should be reported separately. If the intervention under consideration influences life expectancy, and consequently the medical costs in the gained life years are relevant, two separate scenarios must be reported. Then, one scenario presents all costs, including the non-related medical costs in the life years gained; the second can bring to light the effect of including the non-related medical costs in life years gained. It is often difficult, however, to distinguish between related and non-related costs. In that case a scenario including only the intervention costs will be more informative.

4.1.3 Reporting of quality of life
The following characteristics should be reported separately for each quality of life valuation: the questionnaire or valuation method; nationality of the respondents and the perspective: patient valuations or societal valuations. Any deviation from the reference-case analysis (EQ-5D-5L, Dutch value set[29]) should be substantiated. The reporting then should explain why quality of life data from other countries are generalizable to the Netherlands.

If additional measures are presented alongside the standard instrument (EQ-5D-5L), the validity of each must be underpinned on the basis of scientific publications. In a model study using health states the valuations for each health state should be presented separately. If QALYs are estimated on the basis of an 'area under the curve' analysis, the course of the mean quality of life over time should be presented in a table of graph. If the valuations of quality of life are derived from the literature, the following characteristics should be reported separately for each valuation: the questionnaire or valuation method; nationality of the respondents and the perspective taken: patient perspective or societal perspective.

4.1.4 Reporting of missing data
In each economic evaluation based on an empirical study it should be clear in the reporting whether there were any missing data, and if so, how these were accounted for in the analyses. In addition, it is important to report the amount of missing data, whether patients with missing data differ from patients without missing data and whether the issue of missing data was addressed in the study design. In the event that a great amount of data is missing, it is recommended to adjust for this in a number of different ways, so as to explore how sensitive the outcomes are to the method of analysis (see paragraph 2.3.1 for more information).

In an economic evaluation based on a decision model the requested information described above should be reported for all parameters of the model. This will not always be possible in the case of estimates based on published studies; then it is recommended to state explicitly that the study in question does not provide details about any missing data. If an expert panel is used to address missing data the arguments and the methods should be clearly described.

4.2   Reporting of results base-case analysis
The reporting should present the results of the required analyses. In this respect, a distinction is made between the results of the base-case analysis and the results from uncertainty analyses.

The reporting should include both the absolute costs and absolute health effects for the intervention and the comparative treatment, the incremental costs and effects, and the ICER. These results should be presented both with and without discounting. The absolute and incremental costs are presented as total sums, and additionally broken down by cost category.

With regard to the health outcomes, the reporting must minimally include QALYs and the amount of life years gained. If relevant, other additional health outcomes as well as side-effects can be presented separately.

Both an ICER with the QALY as outcome and an ICER with life years gained (LYG) as outcome should be presented. ICERs could also be presented for additional relevant health outcomes.

Dependent on the research question and the intended user, results may be presented from more than one perspective. Note that the results should be presented at least from the societal perspective. If the research question involves subgroups, the results for the separate subgroups should be presented.

In the event that more than two interventions are compared, the results and ICERs of the different interventions should be presented in a well-structured manner. They need to be sorted by increasing costs. In addition, the interventions should be compared by means of a 'full incremental comparison'. The total or absolute costs and effects of the different interventions should be presented graphically as a cost-effectiveness plane (CE-plane). Furthermore, the occurrence of 'extended dominance', if any, must be reported.

4.3   Reporting of uncertainty analyses
The objective of the uncertainty analyses mentioned in paragraph 2.2 is to demonstrate (the influence of) various uncertainties. The reporting of these analyses should at least include the elements mentioned below.

4.3.1   Sensitivity and scenario analyses
The effects of variation of the input parameters in the sensitivity analysis should be presented in a well-structured way. The reporting of the sensitivity analyses for parameter uncertainty should include the minimal and maximal ICER (discounted) of the chosen range, as well as the incremental costs and effects that served as a basis for these ICERs. The upper and lower values of the range must be clearly specified, as well as the reasons why these were applied. It is recommended to present the results of the most influential parameters both in tabular form and graphically by means of a Tornado diagram. Results are always presented together with and compared with results of the base-case analysis.

The results of scenario analyses are preferably presented in tabular form listing the incremental costs and effects as well as the ICER.

4.3.2   *Probabilistic sensitivity analyses and measurement uncertainties*
The results of the PSA or of the bootstrap analysis (in the case of an empiric economic evaluation) on the measurement uncertainty should be presented graphically in a scatter plot showing the incremental costs on the y-axis and the incremental effects on the x-axis (CE-plane).

Results of the PSA should additionally be presented as a cost-effectiveness acceptability curve (CEAC). In general a CEAC reflects the uncertainty around an ICER more transparently than does a confidence interval, because the latter cannot be interpreted if outcomes are distributed over multiple quadrants of the CE-plane. If more than two interventions are being compared, it is advised to present a family of CEACs with a cost efficiency frontier rather than a single CEAC.[30] A possible alternative approach involves calculating the net benefit for each intervention and comparing these benefits with each other. This approach should include a range of ICER reference values (thresholds) that are considered relevant. More information on this topic can be found in the book by Briggs et al.[10]

4.3.3   *Value of information analysis*
The results of EVPI analysis for different reference values of the ICER should be presented in a graph. For more information please refer to the 'Uncertainty and Value of Information' module.

5   Chapter 5 Budget impact analysis

A budget impact analysis (BIA) is an ideal method to explore the financial consequences of (i) acceptation and dissemination of new or optimized healthcare interventions and/or (ii) de-implementation of existing healthcare interventions. The outcome allows a budget holder to make a policy decision regarding the affordability and the required capacity. The BIA takes into account all relevant actual expenses incurred by the paying party, albeit the effectiveness of the intervention is not explicitly analyzed.

5.1   Analytic framework

The outcome of a BIA is the difference in expected expenditure between a scenario in which a new or optimized intervention is accepted and disseminated on the one hand, and a reference scenario in which this is not the case on the other. In most of the situations the standard healthcare in the Netherlands serves as the reference scenario. The expenditure is estimated taking into account changes in prevalence and incidence (in the case of prevention), the size of the treated population, the content of the healthcare interventions offered and in the prices. See the framework in Figure 1.
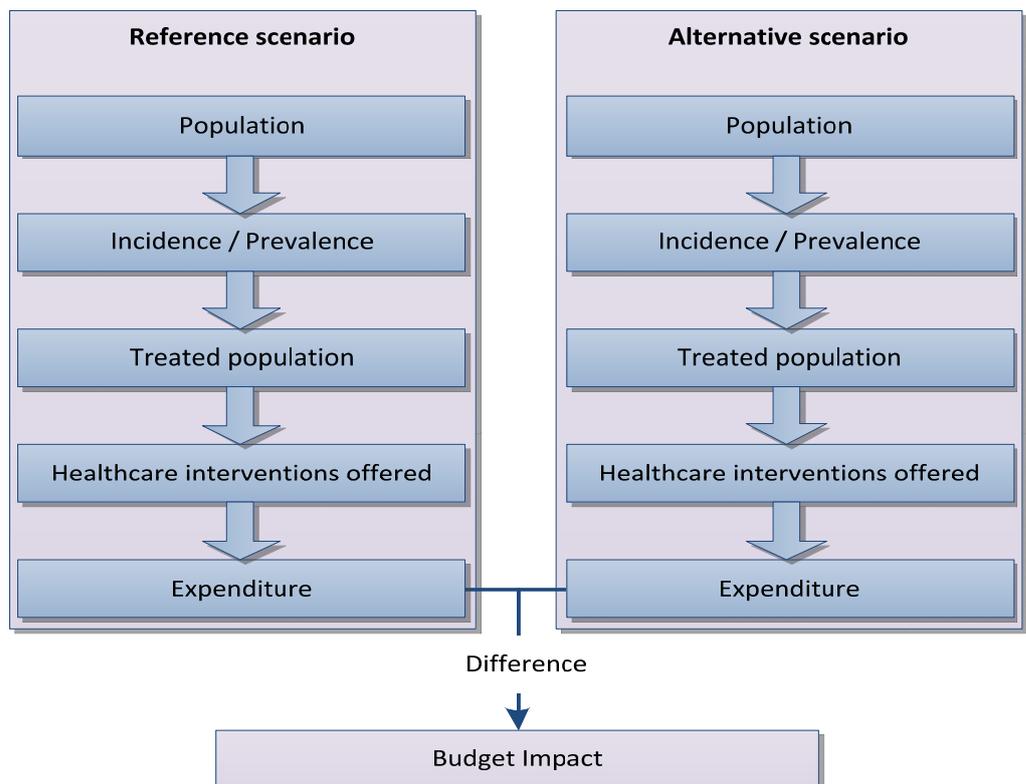

Figure 1. Budget impact analysis framework (adapted from Brosa et al. 2005[31])

The BIA should be designed, conducted and reported in accordance with the internationally accepted principles of ISPOR.[32,33] The following guidelines are derived from these ISPOR principles.

5.2   Perspective

A BIA is performed from the perspective of the budget holder (or: the paying party) who will have to bear the financial consequences of the new or optimized healthcare intervention or the abandonment of existing healthcare.

If the budget holder is at the national governmental level, the BIA should be performed from the societal perspective and include both a broad social quantification of budget effects for society in the broadest sense, in line with economic evaluations, and a calculation of the effect of changes in the net healthcare budgetary framework (BKZ). This framework is largely made up of the Healthcare Insurance Act (Zorgverzekeringswet), the Long-term Care Act (Wet Langdurige Zorg, WLZ) and the Social Support Act (Wet maatschappelijke ondersteuning, Wmo). Supplementary insurance plans fall outside the scope of the BKZ. The net BKZ equals the BKZ excluding the statutory personal contributions of healthcare consumers.

Additional perspectives may be chosen with the approval of the commissioning party.

5.3   Time horizon

The time horizon of BIA should be chosen such that it permits obtaining a result that is useful for the budget holder; it should therefore be chosen in joint agreement with the budget holder. Usually a horizon on the short to medium term is chosen. A 3 years' time horizon is the minimum for the state's national projections, in which the results are presented per year and in addition an indication of the expected market penetration (proportion of the population receiving the new intervention) is given for the years in question. A longer period is needed if the expenditure is likely to undergo changes over the years (for example treatment of chronic diseases, savings on the long term).

5.4   Dynamics of implementation in practice

The framing of the BIA should reflect the complexity and dynamics encountered in real practice as well as the specific features of the Dutch healthcare system. Examples are the influence of personal contributions of healthcare consumers, the rapidity of implementation of new interventions, substitutions of interventions, catch-up effects and the spending on side-effects, complications and if available, therapy adherence. Expenses related to off-label use should also be included in the BIA.[3]

5.5   Items of expenditure, volume measurement and valuation

For information on items of expenditure, volume measurement and valuation, where possible, please refer to the in-depth module 'Costing guidance'.[27] The way in which volumes are measured depends on the perspective of the budget holder. Within the context of the BKZ this involves the tariffs claimed against the insurance funds.

5.6   Reporting

Expenditure should be presented per budget period (typically one year) at the price level of the budget period in question. After all, budgets are commonly presented as nominal figures.

To enable reproduction of the analysis, the final reporting should contain detailed information on the parameter values (volumes and prices) and the calculations. The presentation format must be tailored to the budget holder's information need. Uncertainty analyses for the major input parameters and assumptions should be reported as well.

6   Chapter 6 Specific considerations for various fields of application

This Guideline starts with setting out the reference case, which should be applied to all healthcare economic evaluations. However, this reference case analysis was framed in particular for economic evaluations in the fields of pharmaceutical care and curative care. To evaluate whether the Guideline can be made applicable to other fields as well, such as prevention, diagnostics, medical devices, long-term care and forensic care, this chapter's main objective is to explore whether specific adaptations would then be needed.

The sections 6.1 up to and including 6.5 elaborate on the specific considerations for each of the above-mentioned fields and describe the possible deviations from the reference case analysis. Table 4 provides an overview of the considerations for each of these fields. Any deviations from the reference case analysis should be substantiated for each field. The results of alternative analyses are presented next to those of the reference case analysis.

6.1   Prevention

In an economic evaluation of preventive interventions, it is best to follow the reference case analysis as strictly as possible for the sake of uniformity and comparability. Notably, the framing of the PICOT requires great attention in light of the diversity in preventive interventions. Examples include new legislation, mass media campaigns, disease screening and vaccination. In many cases it is crucial to include intersectoral effects and costs, for example those related to school dropout and judicial involvement, which often are of relevance in preventive interventions. Furthermore, outcomes are often reported as intermediate outcomes, which means that a well-grounded approach for the extrapolation to endpoints is important.

Furthermore, an uncertainty analysis with different discount rates must be performed. After all, in preventive interventions costs are often incurred mainly at the start, whereas the effects will only manifest in the longer term. We would suggest using the discount rates applied in social cost-benefit analyses (MKBA).

Seeing that the outcomes will not be known until on the longer term, a modelling approach is usually required. Specific points of considerations here are the right choice of the model and the choices and assumptions included in the model.

It is not always clear how long the effect of the intervention will be sustained after the intervention has stopped or how long the effect of new legislation will be evident. Habituation may occur for certain interventions, such as an increase in excise duty, leading to a weakening of the effect over the course of time. This waning effect should be accounted for in the economic evaluation.

The selection of the sources for input data is quite essential because often different secondary sources are needed. For example, a wide range of input data may be needed regarding the multiple conditions associated with a single specific risk factor, such as excessive weight. This is the reason why different secondary sources are required.

In a BIA of a preventive intervention the BKZ may be less relevant if large parts of the costs and the effects fall outside the BKZ. Interventions can also be funded from, for example, the budget of the ministry of Education, Culture and Science or

the State budget. If a short 3 years' time horizon is taken for primary prevention, realistic assumptions must be made about the rapidity of the implementation or intervention rollout. This is an issue in for example mass media campaigns or vaccinations.

6.2 Diagnostics

In an economic evaluation of a diagnostic method, it is best to follow the reference case analysis as strictly as possible for the sake of uniformity and comparability. The psychometric properties of the diagnostic tests should be taken into consideration, i.e. the sensitivity and specificity. These properties also determine in part when and for which patients a specific diagnostic test is applied. This means that the patient population and the associated incidence and prevalence should be well defined in advance. In this context, consideration should be given to the synthesis of the diagnostic accuracy of the test, which can be determined with the QUADAS-2 checklist.[34,35] In addition, it is important to explore the consequences of a wrong diagnosis (both false negative and false positive).

The question where the diagnostic test is placed in the test- and treatment strategy is essential in view of the influence of the test results on the subsequent treatment strategy. It should be kept in mind that different test results may lead to different treatment strategies.

Some new diagnostic methods are aimed at ease of use. Diagnostic methods can be appreciated in more ways than only on the basis of the medical outcomes as established with the EQ-5D. Other value components may be identified by directly consulting patients and users with approaches such as a discrete-choice experiment (DCE) and multi-criteria decision analysis (MCDA).[36]

More information about the specific considerations for the field of diagnostics is available in the manual of the NICE Diagnostics Assessment Programme.[37]

6.3 Medical devices

In an economic evaluation of medical devices, it is best to follow the reference case analysis as strictly as possible for the sake of uniformity and comparability. Note however that there is an important shift of emphasis here. In comparison to pharmaceuticals, in the case of devices a stepwise innovation is the rule rather than the exception. This is why a lifelong time horizon would lose much of its impact in the evaluation of devices and should not be automatically chosen. A shorter time horizon may be more relevant but should be well justified.[38,39]

Furthermore, medical devices deviate from other interventions in several essential aspects:
• devices encompass a large heterogeneity of products, ranging from direct patient-related devices to hospital equipment;
• the results of the use of devices may strongly depend on the caregiver's experience with the specific device;
• certain devices are specifically targeted at greater ease of use, which represents a value component independent of the clinical outcomes.

The diversity of devices is enormous, and in the 'healthcare market' we can distinguish two target groups of primary users of diagnostic or therapeutic devices. With regard to application on an individual level, for example the thermometer with which diabetes patient monitor the foot temperature to prevent foot ulcers, there is not much difference with a common intervention; the market consists of the number of patients. With regards to equipment, such as a CT-scanner, the situation is less

clear-cut. In that case, the first layer in the healthcare market is made up of the caregivers who will be using the device. The second layer consists of the patients on whom the caregivers will be applying the device. These two layers should be taken into account when determining the price per unit, also considering that the depreciation charges of the equipment via the procedure are influenced by the utilization degree, or in other terms, the non-optimal use of the device. That said, collective devices are often suitable for use in multiple patient populations, which may have a cost-reducing effect on the price per procedure.

Apart from clinical effectiveness, also the technical effectiveness and technical efficacy of devices may be relevant (intermediate) end points. Examples include diagnostic features, ease of use for the caregiver, or shorter duration of the procedure. Some innovations, such as improved diagnostic performance, may also be expected to have an effect on the clinical outcome; for other innovations such as a shorter duration of the procedure this would seem less obvious.

Due to the use of outcome measures other than quality of life for economic evaluations of medical devices, it is not always possible to perform a CUA. Instead, a CEA is a good alternative.

For some devices, such as implants, the effects strongly depend on the user's experience and training (caregiver or patient). This is notably relevant when extrapolating effectiveness from a controlled study (with experienced users) to the general population.

Other devices, especially body-worn devices, aim to provide ease of use. For this reason the value of devices can be appreciated in more ways than only on the basis of the medical outcomes as established with the EQ-5D. Other value components may be identified by directly consulting patients and users with approaches such as a discrete-choice experiment (DCE) and multi-criteria decision analysis (MCDA).

More information about specific considerations for medical devices can be found in: Food and Drug Administration (FDA) – Medical Devices[40] and Ackerman et al., 2011[41]

6.4 Long-term care

In an economic evaluation of interventions in long-term care, it is best to follow at minimum the reference case analysis for the sake of uniformity and comparability. An important addition to the reference case analysis concerns the objective of the intervention: in many cases the objective is not primarily a health benefit in terms of quality of life assessed with an instrument such as the EQ-5D or life expectancy. Instead, goals are set that mainly target 'well-being' or 'welfare', such as an improved living environment, living independently for as long as possible, or more social interaction.
Solid arguments should be provided for the added elements in the evaluation. It is recommended then to use, next to the EQ-5D, a validated well-being questionnaire for which reference values are available, and which results in a usable, defined utility value (on the basis of tariffs). For the sake of standardization the ICECAP is recommended because this questionnaire is the best-validated one in the Dutch linguistic region.[42]

As the focus here is on well-being or other outcome measures than quality of life in economic evaluations in long-term care, it is not always possible to perform a CUA. Instead, a CEA is a good alternative.

6.5 **Forensics**

In economic evaluations of forensic interventions, it is best to follow the reference case analysis for the sake of uniformity and comparability. An important addition to the reference case analysis concerns the objective of the intervention. Many interventions in this domain aim to achieve improvement of behavior, for example reducing criminal behavior or antisocial behavior. Interventions of this type may not only target the mental health of the 'patient' (assessed with quality of life measures), but also be aimed at modifying the environment, whereby behavioral improvement is achieved but no 'health' improvement. Furthermore, the effects of the interventions are often multi-dimensional and tend to transcend health issues. In addition to recidivism reduction, improvement of school performance or family functioning, reducing the number of contacts with the criminal justice system, causing less nuisance or cutting back on narcotics use may be other important targets. Note that in this context it may be problematic to precisely delineate the boundaries between mental healthcare, forensic care and preventive care.

Solid arguments should be provided for the added elements in the evaluation. As long as broader outcome measures suited for economic evaluations in this domain are lacking, it is crucial to select an unambiguous outcome measure that fits well with the primary objective of the intervention. Thus, the outcome measure 'criminal activity free years' (CAFYs), in which time and criminality are taken into account, is an option for interventions targeting criminality. For interventions targeted at drug abuse, the outcome 'drug abuse free years' (DAFYs) is worth considering. Costs and effects can be measured both for the 'patients' and the 'victims', but please note that regarding the latter group the number of 'victims' must be carefully estimated.

Table 4: Specific considerations per field of application

	REFERENCE CASE	PREVENTION	DIAGNOSTICS	MEDICAL DEVICES	LONG-TERM CARE	FORENSICS
Perspective	Societal perspective					
PICOT	P: Population in the Dutch setting	Extra attention needed	Extra attention needed	Distinction between collective devices and individual devices		
	C: Standard and/or usual care in the Netherlands	Extra attention needed; diversity in interventions	Clarity about diagnostic accuracy			
	T: Expected lifetime	Not always expected lifetime		Not always expected lifetime		
	Scoping beforehand with relevant stakeholders					
Type of economic evaluation	Cost-utility analysis (CUA)			Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis (CEA)
Discounting	Costs 4% and effects 1.5%	Different discount rates				Different discount rates
Uncertainty and sensitivity analyses	Univariate, probabilistic sensitivity analysis and scenarios			Level of extrapolation: for certain devices, such as implanted ones, the effects strongly depend on the user's experience and training (surgeon or patient).		
Effectiveness	Model-based evaluation: Systematic review					
Costs	All costs in the healthcare system, patient/family, and other sectors. Productivity losses: friction cost method.	Extra attention needed for intersectoral costs		Utilization degree influences the fixed price per unit and thus the cost price (such as depreciation		Quantify costs for 'patient' and 'victim'

Table 4: Specific considerations per field of application

	REFERENCE CASE	PREVENTION	DIAGNOSTICS	MEDICAL DEVICES	LONG-TERM CARE	FORENSICS
				charges). Collective devices: employable in more than one population has cost-reducing effect on the price per procedure (synergy gains)		
Effects	Expressed in QALYs, at minimum the EQ-5D-5L			Technical effectiveness and technical efficiency	Welfare/ well-being	Welfare/ well-being CAFYs/ DAFYs
Reporting of results	Total costs and effects, incremental costs and effects, incremental cost-effectiveness ratio					
	Univariate sensitivity analysis: Tornado diagram and table Scenario analysis: table Probabilistic sensitivity analysis: CE-plane and CEAC					

References

[1] Centre for evidence based medicine: Using the PICO to define the research question. http://www.cebm.net/finding-the-evidence-1-using-pico-to-formulate-a-search-question/

[2] Rapport Beoordeling stand van de wetenschap en praktijk (geactualiseerde versie 2015): Zie website zorginstituut Nederland: www.zorginstituutnederland.nl

[3] EUnetHTA. EUnetHTA guidelines: www.eunethta.eu/eunethta-guidelines.

[4] Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, Sullivan SD. Key principles for the improved conduct of health technology assessments for resource allocation decisions. International Journal of Technology Assessment in Health Care. 2008; 24(3): 244-258.

[5] Ramsey SD, Wilke RJ, Glick H, Reed SD, Augustovski F, Jönsson B, et al. Cost-effectiveness analyses alongside clinical trials II – an ISPOR Good Research Practices Task Force Report. Value in Health. 2015; 18 (2): 161-172.

[6] Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices – Overview: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force 1. Value in Health. 2012: 15: 796-803.

[7] Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modelling Good Research Practices Task Force-6. Value in Health. 2012; 15: 835-842.

[8] Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Fourth edition 2015. Oxford University Press.

[9] Bilcke J, Beutels P, Brisson M, Jit M. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Medical decision making. 2011; 675-692.

[10] Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. 2006. Oxford University Press

[11] Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. Second Edition 2014. Oxford University Press.

[12] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. British Medical Journal. 2009; 338:b2393

[13] Oostenbrink JB, Al MJ. The analysis of incomplete cost data due to dropout. Health Economics. 2005; 14:763-776.

[14] National Institute for Health and Clinical Excellence Decision Support Unit. Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data, 2013. Toegankelijk via: www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf

[15] Vemer P, Corro Ramos I, Van Voorn GAK, Al MJ, Feenstra FL and the AdViSHE Study Group. AdViSHE: A validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. Accepted for publication.

[16] Schünemann H, Brożek J, Guyatt G et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013.

[17] Strengtening the Reporting of Observational Studies in Epidemiology (STROBE). STROBE statement – checklist of items that should be included in reports of cohort studies. 2009 Toegankelijk via: strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_v4_cohort.pdf

[18] Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials

unnecessary? Picking signal from noise. BMJ. 2007;334:349-51.

[19] Schünemann H, Hill S, Guyatt G, et al. The GRADE approach and Bradford Hill's criteria for causation. Journal of Epidemiology & Community Health. 2011;65:392-395.

[20] Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis. http://www.ohri.ca/programs/clinical_epidemiology/nosgen.pdf

[21] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology & Community Health. 1998; 52:377–384. http://jech.bmj.com/content/52/6/377.full.pdf+html

[22] Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. Medical Decision Making. 2013;33(5):597-606. till Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Medical Decision Making. 2013;33(5):671-8.

[23] Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. British Medical Journal. 2009;338:b1147.

[24] Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014; 3;9(7):e99682

[25] Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 5.0.2 updated September 2009. http://www.cochrane-handbook.org/

[26] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappeleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. Value in Health. 2011; 14(4): 417-428.

[27] Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan SS. Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. In opdracht van Zorginstituut Nederland. Geactualiseerde versie 2015.

[28] Maastricht University, 2014. Handleiding intersectorale kosten en baten van (preventieve) interventies: www.maastrichtuniversity.nl/web/Main/Sitewide /Content/ClassificatieIdentificatieEnWaarderingVanIntersectoraleKostenEnBatenVanP reventieveInterventies.htm

[29] Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the 5 level version of EQ-5D. Value in Health. Accepted for publication.

[30] Fenwick E, O'Brien B, Briggs A. Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions. Health Economics. 2004; 13: 405-415.

[31] Brosa M, Gisbert R, Rodríguez Barrios JM, et al. Métodos y aplicaciones del análisis del impacto presupuestario en sanidad. Pharmacoeconomics Spanish Research Articles 2005;2(2):65–79.

[32] Sullivan SD, Mauskopf JA, Augustovski F, et al. Budget impact analysis. Principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health. 2014;17(1): 5-14.

[33] Goettsch WG, Enzing JJ. Review: Report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health. 2014;17(1):1-2.

[34] University of Bristol. QUADAS - QUADAS-2. Toegankelijk via www.bristol.ac.uk/quadas/quadas-2/

[35] Whiting PF, Rutjes AWS, Westwood ME et al. and the QUADAS-2 group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011; 155: 529–36

[36] Lee DW, Neumann PJ, Rizzo JA. Understanding the Medical and Nonmedical Value of Diagnostic Testing. Value in Health. 2010; 13(2): 310-314.

[37] National Institute for Health and Clinical Excellence, 2011. Diagnostics Assessment Programme Manual:http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf.

[38] Van de Wetering G, Woertman WH, Verbeek AL, Broeders MJ, Adang EM. Quantifying short run cost-effectiveness during a gradual implementation process. European Journal of Health Economics. 2013; 14(6): 911-8.

[39] Van de Wetering G, Woertman WH, Adang EM. A model to correct for short-run inefficiencies in economic evaluations in healthcare. Health Economics. 2012; 21(3): 270-81.

[40] Food and Drug Administration (FDA). Patient Preference Information – Submission, Review in PMAs, HDE Applications, and De Novo Requests and Inclusion in Device Labeling. May 2015. Toegankelijk via: www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM446680.pdf

[41] Ackerman SJ, Dix Smith M, Ehreth J, Eldessouki R, Sullivan E. Therapeutic and diagnostic device outcomes research. Lawrenceville (USA): ISPOR, 2011. www.ispor.org/publications/MDD_books.asp

[42] Makai P, Koopmanschap MA, Brouwer WB, Nieboer AA. A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands. Health and Quality of Life Outcomes. 2013;11:57