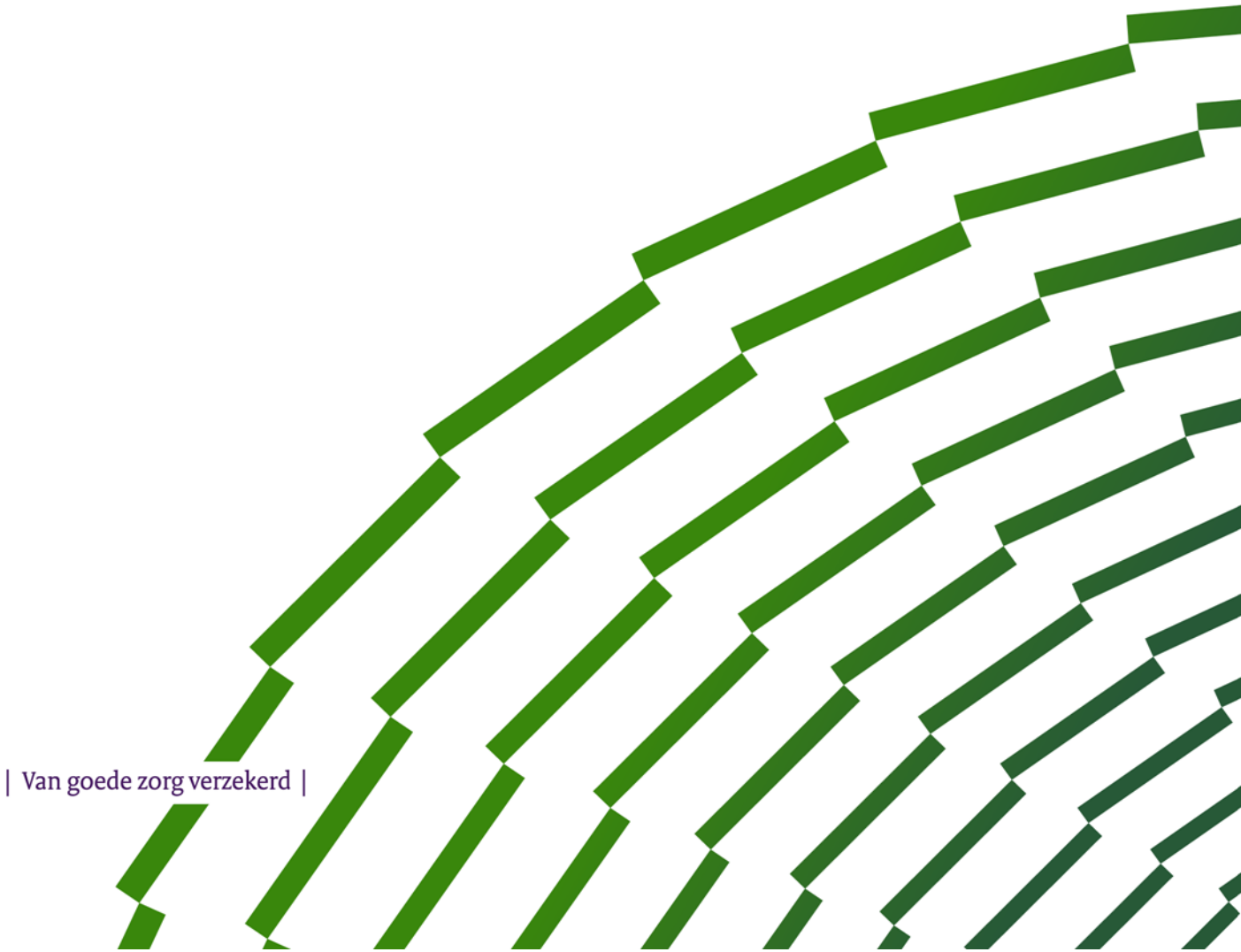
Zorginstituut Nederland

Guideline

for economic evaluations in healthcare

2024 version



| Van goede zorg verzekerd |

Table of contents

Preface

Since the first publication of the Dutch Guidelines for pharmacoeconomic research, twenty-five years ago, both science and the National Health Care Institute have advanced. Updates of those first guidelines were published in 2006 and 2016, based on new scientific insights, for example in relation to reporting uncertainty regarding the outcomes of economic evaluations. Within the National Health Care Institute, cost-effectiveness has become more and more important in reimbursement decisions over the years. Cost-effectiveness is one of the four package criteria. Since the National Health Care Institute is responsible for systematic and transparent considerations relating to the basic healthcare package, it has become increasingly important to standardise economic evaluations. That is why a 'reference case' was introduced in the previous version of the guideline, dating from 2016.

Economic evaluations which comply with the prescribed reference case can be compared with each other in terms of this package criterion. This third update of the guideline defines the reference case even more specifically, for example in relation to the healthcare costs in life years gained, which now form an integral part of the cost-effectiveness analysis for the first time. The choice to conduct the economic evaluation from a societal perspective has been constant over the years, and the guideline committee did not question this decision this time either. In an international context the Netherlands is one of the countries which has traditionally taken outspoken decisions on this because few countries apply such a comprehensive societal perspective. Choosing this perspective was also possible because Dutch scientists perform a great deal of research into measuring and evaluating societal costs and effects so that these can be actually included in the economic evaluation.

The guideline committee was made up of eight members who are all very familiar with the assessment procedures of the National Health Care Institute and were supported by some of its expert staff. Although the guideline is published by the National Health Care Institute, the committee also took into account the fact that the guideline is also being used for economic evaluations outside the remit of the National Health Care Institute. The Netherlands is a country where the field of health economics flourished many years ago and is practised at a high level. Research into the effectiveness of new medical developments is, for example, performed at all the university medical centres, which is often a condition for the awarding of research subsidies by the Dutch organisation for Health Research and Development (ZonMw) or the research programmes of the European Union. The guideline committee's ambition and expectation is that this new guideline will also be used for such economic evaluations.

In the upcoming years, both the National Health Care Institute and Dutch health economists will be able to continue their contributions to the careful substantiation of healthcare policy and, consequently, appropriate care, because of this updated guideline.

On behalf of the committee charged with revising the guideline,

Prof. Ardine de Wit
Chair of the committee charged with revising the guideline

Abbreviations

AdVISHE	Assessment of the Validation Status of Health-Economic decision models
AIC	Akaike information criterion
BIC	Bayesian information criterion
CEAC	Cost-effectiveness acceptability curve
CE-plane	Cost-effectiveness plane
CPI	Consumer price index
CVZ	Health Care Insurance Board (College voor Zorgverzekeringen)
DCE	Discrete choice experiment
ENBS	Expected net benefit of sampling
EQ-5D	EuroQol-5 dimensions
EVPI	Expected value of perfect information
EVPPI	Expected value of partial perfect information
EVSI	Expected value of sample information
GCP	Good clinical practice
GRADE	Grading of recommendations, assessment, development and evaluations
ICER	Incremental cost-effectiveness ratio
IKB	Intersectoral costs and revenues
ISPOR	The International Society for Pharmacoeconomics and Outcomes Research
ITC	Indirect treatment comparison
KEA	Cost-effectiveness analysis (CEA)
KNAW	Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen)
KUA	Cost-utility analysis (CUA)
MAR	Missing at random
MAIC	Matching adjusted indirect comparison
MCAR	Missing completely at random
MCDA	Multi-criteria decision analysis
MKBA	Social cost benefit analysis (SCBA)
MTC	Mixed treatment comparison
NMA	Network meta-analysis
NMB	Net monetary benefit
MNAR	Missing not at random

PAID	Practical Application to Include future Disease costs
PICOTS	Patient, Intervention, Comparison, Outcome, Time, Setting
QALY	Quality adjusted life year
RCT	Randomised controlled trial
SHELF	Sheffield elicitation framework
STC	Simulated treatment comparison
VOI	Value of Information

Introduction

Aim of the guideline

The aim of this guideline is to set a clear framework for those who perform economic evaluations for reimbursement decisions in healthcare. The starting point of drafting this guideline is therefore a decision-based approach with methodological choices being based on the premise that ultimately a decision can be taken within the context of the Dutch healthcare sector.

History

In 1999 the Health Care Insurance Board (College voor Zorgverzekeringen, CVZ) issued the 'Guidelines for pharmacoeconomic research'. This document was revised in 2006 and, since then, has been used as an assessment framework for pharmacoeconomic evaluations which became part of medicines reimbursement files. Based on the 2006 version assessments were carried out to determine whether there was sufficient cost-effectiveness evidence of a medicine for which a reimbursement had been requested. The Guidance on Outcome Research was published in 2008 and the Cost Research Manual (also referred to as the Costing Manual) was published in 2010, meaning that three guidelines were available. In 2016, these documents were combined to create a single new guideline, referred to as the Guideline for Economic Evaluations in Healthcare. In contrast to the three separate guidelines, the 2016 guideline not only focused on medicines, but also on other types of care interventions. The guideline was also amended in line with the latest scientific developments.

Due to continuous methodological developments in the field of economic evaluations, the 2016 guideline needed to be revised again in 2024. **Table 1** shows an overview of the guideline's history.

Table 1: History of the guideline

Year	Title/version	Summary of the most important changes compared to the previous version
1999	Guidelines for pharmacoeconomic research	n/a
2006	Revised version of the Guidelines for pharmacoeconomic research	Update
2008	Guidance on outcome research	n/a
2016	Guideline for economic evaluations in healthcare	- Merging of the various guidelines, namely the Guidelines for pharmacoeconomic research (2006) and the Guidance on outcome research (2008) - Major revision of content - Addition of in-depth modules: – Costing manual* – QALY and quality of life measurements – Uncertainty and value of information analyses
2024	Guideline for economic evaluations in healthcare, revised version 2024	Update

*See the '**Costing manual**' in-depth module for the history of its various revisions.

Changes compared to the 2016 guideline

The most important changes in the current version compared to the previous version of the guideline are as follows:

- The discount rate for costs has been adapted to 3%;
- Medical costs in life years gained are included in the base case analysis;

- The results of both the base case analysis and scenario analyses must be based on the probabilistic analysis;
- Value of information (VOI) analysis, in the form of the expected value of perfect information (EVPI) and expected value of partial perfect information (EVPPI), is an obligatory element in the case of model-based economic evaluations;
- The quality of life of informal caregivers must be included in a scenario analysis when relevant;
- The EQ-5D-Y must be used to measure the quality of life of children aged 8-12.

The guideline also contains greater clarification regarding:
- empirical economic evaluations;
- the extrapolation of time-to-event data;
- subgroup analyses;
- uncertainty analyses;
- validation;
- expert opinion and expert elicitation.

Non-permanent committee

A temporary committee was set up to support the National Health Care Institute with the revision of the guideline. Its members were Prof. G.A. de Wit (chair), Dr H.M. Blommestein, Prof. J.E. Bosmans, Prof. W.B.F. Brouwer, Prof. J.J. van Busschbach, Prof. T.L. Feenstra, Prof. H. Koffijberg and Prof. M.A. Joore. The following staff from the National Health Care Institute were involved in the update: Dr H.A. Geuzinge (project lead/secretary), Dr M. El Alili, Dr J.J. Enzing and Dr S. Knies.

Document structure

Structure of the 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. The main body of the document consists of six chapters. Chapter 1 provides an overview of the requirements of the reference case which all economic evaluations must meet. Chapter 2 examines the framework of an economic evaluation. Chapter 3 addresses the various data which are required for an economic evaluation. Chapter 4 focuses on the method. Chapter 5 contains details on how the results of an evaluation are reported. Chapter 6 deals with a number of specific areas of concern.

In-depth modules

There are three different in-depth modules, namely:
- The costing manual (‘Kostenhandleiding’)
- QALY and quality of life measurements (‘QALY en kwaliteit-van-leven-metingen’ (only available in Dutch))
- Uncertainty and Value of Information analyses (‘Onzekerheid en Value of Information analyses’ (only available in Dutch))

The in-depth modules provide more detailed information on specific elements of the guideline and are an integral part of it.

1 Reference case

In order to increase the extent to which economic evaluations conducted in the Netherlands can be compared, as well as to enhance their quality, this guideline starts with an explanation of the reference case to which all economic evaluations carried out for decision-making have to comply with (see **table 2**). Other choices may be made in scenario analyses, if relevant. Sufficient justification and a clear description must be provided for these choices. Any additional analyses cannot, replace the reference case. **Chapter 6** looks at a variety of topics that may be relevant to the various areas of application.

Table 2: Explanation of the reference case

Element	Reference case	See paragraph
Perspective	Societal perspective	2.2
PICOTS	P: Intended patient population	2.3
	I: The intervention under consideration	
	C: Standard and/or most usual intervention in the Netherlands (if this is not the most cost-effective intervention, a comparison must also be made with the most cost-effective intervention)	
	O: See 'effects' and 'costs' in this table	
	T: A lifetime horizon in the case of model-based evaluations. In empirical evaluations, the time horizon need to be such that a valid statement can be made about the effectiveness and costs	
	S: Context in which the care is delivered in the Netherlands	
Type of economic evaluation	Cost utility analysis	2.4
Effectiveness data	Systematic review/meta-analysis, clinical trial and other suitable evidence	3.1
Costs	All costs within healthcare (also during life years gained), patient and family costs, as well as costs in other sectors. Productivity losses: friction cost method. Use reference prices from the 'costing manual'	3.2
Effects	Expressed in QALYs, determined with the EQ-5D-5L (or the EQ-5D-Y in the case of children aged 8-12) with Dutch valuations	3.3
Discounting	Costs 3% and effects 1.5%	4.2
Subgroup analyses	Subgroup analyses in the case of known heterogeneity in treatment effect or in the baseline risk on the basis of patient or disease characteristics	4.6
Uncertainty analyses	In model-based evaluations: deterministic uncertainty analyses, probabilistic analysis, scenario analyses and value of information analysis In empirical evaluations: deterministic uncertainty analysis, bootstrapping, scenario analyses	4.7
Validation	In model-based evaluations: the conceptual model, input data, the software-based implementation and model outcomes.	4.8

	In empirical evaluations: pre-processing of the data, assumptions relating to uptake of care, the statistical analysis and the outcomes.	
Reporting results	Total costs and effects for the intervention and comparative intervention(s), incremental costs and effects, ICER, NMB	5.3.1
	Deterministic uncertainty analysis: tornado diagram and table Scenario analysis: table Probabilistic of bootstrap analysis: CE-plane and CEAC Value of information (in the case of model-based economic evaluations): patient and population EVP(P)I and EVP(P)I curve	5.3.3 to 5.3.5

ICER: incremental cost-effectiveness ratio, NMB: net monetary benefit, CE-plane: cost-effectiveness plane. CEAC: cost-effectiveness acceptability curve. EVPI: expected value of perfect information. EVPPI: expected value of partial perfect information

2 Framework of the economic evaluation

The economic evaluation of a healthcare intervention must be clearly defined and designed prior the start of the analyses. It is important to note that all the following aspects must be clearly described, and all choices must be justified.

2.1 Objective and user

The first step in an economic evaluation is to define the objective. It is important to define the future user(s) of the evaluation since their decision-making problem should be linked to the objective. Examples of users are the National Health Care Institute, health insurers or health facilities.

2.2 Perspective

Economic evaluations should be conducted and reported from a societal perspective. This perspective includes all relevant societal costs and benefits, irrespective of who bears the costs or to whom the benefits go.

In addition to the societal perspective, scenario analyses can be conducted from other perspectives, such as the healthcare perspective (see **paragraph 3.2**).

2.3 PICOTS

After defining the user and the perspective of the evaluation, the following stage is to formulate the specific research question for the economic evaluation. The evaluation is conducted on the basis of a research question formulated in accordance with the PICOTS method:[1]

- ***P**atient* = the intended patient population the intervention is aimed at.
- ***I**ntervention* = the intervention under consideration.
- ***C**omparison* = the comparative intervention(s).
- ***O**utcome* = the relevant outcomes/outcome measures.
- ***T**ime* = the relevant time span for which effects and costs must be measured.
- ***S**etting* = the context in which the care is delivered, when it can make a difference to the effect of the intervention, for example primary or secondary healthcare.

These elements are clarified in more detail in the following subparagraphs.

2.3.1 Patient (P)

In Dutch clinical practice, the patient population or target population is the group for whom the intervention is intended. To describe the patient population and relevant subgroups, Dutch epidemiological data must be used. Differences in age, gender and disease characteristics may exist between and within populations, influencing the results significantly. **Paragraph 4.6** clarifies when subgroup analyses have to be conducted.

2.3.2 Intervention (I)

The intervention under consideration should be an intervention that can be applied in Dutch clinical practice within the defined patient population. An accurate description is required of how the patient (or caregiver) is going to use the intervention. Moreover, it must be made clear whether the intervention replaces or supplements an existing intervention.

2.3.3 Comparison (C)

The intervention under consideration must be compared with the standard or usual intervention in Dutch practice at the time.[1] If the standard or usual intervention is not the most cost-effective strategy among the alternatives available in the Netherlands at that moment, the intervention that is most cost-effective should also be added for comparison.
The standard intervention is the intervention which is regarded as the first choice in accordance with clinical guidelines or practice.

A comparison with the usual intervention is appropriate if no standard intervention exists. The usual intervention refers to the intervention which is used most frequently in conjunction with

the defined patient population and with which people have the most experience in daily practice. The usual intervention may also imply that no treatment is offered, but rather a recommendation to wait and observe, possibly combined with symptom relief (supportive or palliative care).

In a situation where multiple standard or usual interventions exist, separate comparisons must be made. This can be done by means of a fully incremental analysis (see **paragraph 5.3.1**). If the standard or usual intervention is not the most cost-effective intervention, a comparison will have to be carried out with multiple interventions, using a fully incremental analysis.

2.3.4 Outcomes (O)

The outcomes of a cost-effectiveness analysis are incremental costs and incremental effects of an intervention compared to one or more interventions. Costs are influenced by a variety of factors, including the uptake of care, the use of the interventions, prices, but also by, for example, productivity losses. More information about (types of) costs to be included, as well as the determination of care uptake and prices, can be found in **paragraph 3.2** and the **'costing manual' in-depth module**.

The effects are expressed in QALYs (when applying a cost-utility analysis, see **paragraph 2.4**). QALYs are calculated by multiplying health-related quality of life valuations (often referred to as 'utilities') by life years. Quality of life is a key patient-oriented outcome in economic evaluations. It is measured using the generic EQ-5D-5L measurement tool (see **paragraph 3.3** and the **in-depth module 'QALY en kwaliteit-van-leven-metingen'** (only available in Dutch)).

2.3.5 Time (T)

The time horizon of an economic evaluation must be long enough to cover all relevant costs and effects. A model-based economic evaluation is frequently undertaken in the case of life-prolonging care interventions, then the time horizon is lifelong. If a lifelong time horizon is irrelevant (for example in the context of a short-term illness), a shorter time horizon could be used. In the case of empirical economic evaluations, the time horizon will often be shorter than lifelong.

2.3.6 Setting (S)

It is important to consider the implementation of the intervention and the comparator intervention(s) in Dutch practice. The organisational setting and the characteristics of care providers and patients are important in this respect. This is because the outcomes (in terms of costs and effects) of care interventions might vary depending on the situation. For example, the treatment line or the location where the intervention is used (at home or in a care facility), can play a role. Additionally, different usual interventions can be used in various treatment facilities. Examples of care provider characteristics include for instance specialization and the amount of experience with the intervention (see also **paragraph 6.3** about learning effects).

2.4 Analysis technique

There are various types of economic evaluations. Cost-effectiveness analysis (CEA) and cost utility analysis (CUA) are most commonly used. However, in daily practice these terms are often used interchangeably. In both cases (CEA and CUA), the difference in costs (the incremental costs) is compared to the difference in effects (the incremental effects) of the interventions under consideration. A CEA uses (clinical) effect sizes (for example blood pressure, event free survival, or life years gained). A CUA expresses effects in QALYs, by multiplying quality of life values with the life years. A CUA has the advantage of allowing for comparability of the outcomes of analyses across different patient populations and care interventions. A CUA should be performed as standard if carrying out an economic evaluation for a reimbursement decision (see **Chapter 1**).

The economic evaluation can be performed as part of an empirical economic study, or by using a simulation model. More information about empirical economic evaluations and model-based economic evaluations can be found in **paragraph 4.1**.

3 Data

3.1 Effectiveness

If data from the literature is required with regard to the relative effectiveness of an intervention, a systematic literature review is obligatory. On the basis of studies identified, an indirect comparison can be performed if necessary.

3.1.1 Systematic literature review

In accordance with a predefined protocol, the literature review entails a systematic search for available relevant clinical trials that align as much as feasible with the PICOTS. The search technique, relevant databases and time constraints must be described (in an appendix) to ensure the review is replicable. An overview of all selected studies and the applied selection criteria must be provided. Moreover, the reason why studies have been included or excluded for the economic evaluation should be reported. The literature review is not limited to a certain study design. The effectiveness data must be based on studies which provide the greatest confidence about the effects in the described patient population.[1]

3.1.2 Study designs

If randomised controlled trials (RCTs) are available, these are preferably used as a starting point. It is important that the RCTs are of good quality and in alignment with the PICOTS. A well conducted RCT reduces the possibility of bias and can thus provide the maximum degree of certainty about the causality between the intervention and the observed effect.[1] An RCT should preferably be carried out on a (three) double-blind basis (if possible and desirable, which is often the case with pharmaceuticals).

Observational studies can also be used to provide evidence for the economic evaluation. However, the greater chance of bias (for example confounding, information and selection bias) must be considered, which is different from studies with an experimental design. On the other hand, an observational study will have better external validity than an RCT. See **paragraph 4.4** for methods to adjust for bias.

Sometimes, a new intervention has only been investigated in a single-arm study. This type of study involves assigning the new intervention to all participants. Determining the relative effectiveness based simply on a single-arm study is impossible. Therefore, these studies can only be used in combination with a control arm from an external cohort in an indirect comparison. This form of indirect comparison often implies significant uncertainty and is hence never the preferred approach.

3.1.3 Indirect comparisons

In an indirect comparison the outcomes of study arms from different studies are compared. These are preferably study arms from randomised studies for an anchored comparison. Non-randomised study data, such as observational cohorts and single-arm studies, may also be used. When performing indirect comparisons it is important that studies are carefully selected in order to reduce the possibility of bias.

An indirect comparison can be performed using a Bayesian[2-7] or frequentist approach.[8] Using a network meta-analysis (NMA) (an indirect treatment comparison (ITC) or a mixed treatment comparison (MTC)) is preferred.[9-12] However, it is not always feasible to perform an NMA, for example when there is no literature available for the essential comparisons (incomplete evidence network) or when there is too much heterogeneity between studies. Additionally, problems may arise when too many interim comparisons are necessary in an MTC network. An indirect comparison can also be performed by means of a simulated treatment comparison (STC) or a matching adjusted indirect comparison (MAIC).[12] The relevant conditions and assumptions must be clearly described and justified for each method.

3.2 Costs

Economic evaluations should be conducted from a societal perspective, which indicates that all relevant costs and effects are part of the analysis, irrespective of who bears them. Relevant costs are those incurred in healthcare (for example hospital admissions and healthcare costs in life years gained), costs for the patient and the family (for example informal care) and costs incurred in other social sectors (such as productivity losses, education or the judicial system).

Costs are calculated using the following formula: costs = volume*price. Units should be measured to determine the volume component, and the price per unit can be used to value the units in order to calculate the total cost. The methods described in the **'costing manual' in-depth module** must be used for the identification, measurement and valuation of costs. This will ensure consistency and standardisation of cost identification, measurement and valuation in economic evaluations of care interventions.

3.2.1 Types of costs and volume measurement

Table 3 shows an overview of the relevant cost categories associated with the societal perspective. All cost categories must be described and explained separately.

When it comes to quantifying volumes (quantities), different methodologies apply to each cost category. Measuring requires determining the deployment of people and resources during, after and in some cases also before the intervention. Ideally, Dutch sources should be used, due to the great variety in care uptake between countries. Volume data from international studies must be validated for the Netherlands if they are used. In addition, it is important to ensure that no double counting occurs between different cost categories.

Table 3: Perspectives and associated cost categories

Perspective	Cost categories	Specification of costs
Societal perspective	Costs in healthcare	- All healthcare costs directly related to the condition - All healthcare costs incurred during life years gained
	Costs for patients and family	For example: - Travel expenses - Own contributions - Time-related costs - Costs of informal care
	Costs in other sectors	Costs incurred in sectors outside the healthcare sector, for example by municipalities, in education, or by volunteers Productivity losses: costs of absenteeism or unproductivity during paid work (presenteeism) and unpaid work
Healthcare perspective	Costs in healthcare	- All healthcare costs directly resulting from the intervention - All healthcare costs incurred during life years gained

Source: Drummond et al. (2015)[13]

3.2.1.1 Costs in healthcare

The 'costs in healthcare' category includes all healthcare costs associated with or influenced by the condition under consideration. In this context, all costs incurred during the entire time horizon of the economic evaluation are included. The costs of treating any side effects or complications also belong in this cost category.

Healthcare costs incurred as a result of life years gained should also be considered in the base case analysis. This includes all healthcare costs incurred as a consequence of an intervention's life-prolonging effect. These costs can be categorized as 'related costs' (such as blood thinners after a heart procedure) or 'unrelated costs' (such as the costs of a broken hip in the life years gained after a heart transplantation). The Practical Application to Include future Disease costs (PAID) can be used to determine these costs.[14]

3.2.1.2 Costs for patients and family

Costs for patients and family arise outside formal healthcare but have a direct relationship with the disease or intervention. These costs may consist of, among other things, travel expenses and personal payments, as well as time-related costs.

If the number of visits to a care facility or care provider is influenced by an intervention, these must be included in the economic evaluation. Because it is generally registered, the frequency of visits to a care institution or care provider can often be accurately determined. Please refer to the **'costing manual' in-depth module** for the average distance travelled by patients to a care institution and the monetary valuation of travel expenses.

One of the major cost items in this context concerns the time spent in the provision of informal care. Informal care is defined as unpaid and often long-term care for sick family members or friends. This could include caring for someone, or assisting with daily activities. Providing informal care may limit the time available for paid work, unpaid work and leisure time. The iMTA Valuation of Informal Care Questionnaire (iVICQ) can be used to determine how many hours someone has spent on informal care.[15] The monetary valuation of informal care hours should be based on the replacement costs of domestic care. More information on the valuation of informal care can be found in the **'costing manual' in-depth module.**

3.2.1.3 Costs in other sectors

Costs incurred in other sectors are highly dependent on the intervention to be evaluated and the patient population concerned. For example, costs for special education and the judicial system could be relevant. If cost differences in these areas are expected between the compared interventions, these must be included in the analysis. More information on the valuation of costs within other sectors is available in the **'costing manual' in-depth module**. More information can also be found in the 'Manual of intersectoral costs and benefits'.[16] The results of the PECUNIA project provide examples of the method to be considered for the evaluation of costs within other sectors.[17] The 'guide to costs-benefits analysis in the social domain' also offers an analytical approach.[18]

Interventions aimed (also) at the working population can influence work-related productivity. Costs will be incurred if the disease or the intervention prevents people from being productive at work. Such costs are part of the reference case. Productivity related to both paid and unpaid work is equally important in this respect. In both cases the productivity losses (or gains) involved must be specified and valued. In the case of paid work, there is a distinction between reduced productivity at work ('presenteeism') and absence from work (absenteeism). Both may be important. One frequently used instrument to measure productivity losses is the iMTA Productivity Cost Questionnaire (iPCQ).[19]

The friction cost method must be used for valuation of productivity losses. The starting point in this context is that sick personnel who are absent for an extended period of time can be replaced. Productivity losses will then primarily occur during the time an employer needs to replace a sick employee, the friction period. More information and clarification of the method for calculating productivity losses using the friction cost method is available in the **'costing manual' in-depth module.**

3.2.2 Valuation

The next step is to value the identified and measured volumes in monetary units. The **'costing manual' in-depth module** offers reference prices for a number of common units, in addition to guidelines for conducting a cost price research. If specific required costs have not been reported

in the 'costing manual', alternative sources can be used (see the **'costing manual' in-depth module**).

3.2.3 Adjustment for inflation/price indexation

When determining costs in economic evaluations it is essential to specify the year to which the prices apply. If cost prices of different units or volumes do not come from the same calendar year, the prices should be adjusted for inflation based on the same basic year. This is done using the Dutch consumer price index (CPI), which can be found in the Statistics Netherlands (CBS) StatLine database. The **'costing manual' in-depth module** contains further information and an explanation.

3.3 Quality of life

In a CUA, effects are expressed in QALYs by multiplying health-related quality of life (also referred to as 'utilities') by life years. This paragraph describes how quality of life should be measured. More information on this can be found in the **in-depth module 'QALY en kwaliteit-van-leven-metingen'** (only available in Dutch).

3.3.1 Measuring quality of life

Validated, generic quality of life questionnaires completed by patients should be used to determine their quality of life. The validated questionnaires, in combination with a so-called valuation set, allow for the evaluation of a patient's health status. A health status valuation is also referred to as a 'utility'. The valuation rates must be determined through sampling among the general public. Thus, the patient describes his/her health status using a validated questionnaire, and the resulting health status is graded using a fixed rate based on preferences among the general population. These steps are explained in more detail in the **in-depth module 'QALY en kwaliteit-van-leven-metingen'** (only available in Dutch).

A wide variety of questionnaires exist for measuring quality of life, but rates based on preferences among the general population are scarce. Given the potential discrepancies in the valuation of quality of life due to the use of different questionnaires, quality of life should be measured with the EQ-5D-5L and valued using the Dutch tariff.[20] The EQ-5D-Y-3L questionnaire is available for children aged 8 to 12 years.[21] For children under the age of 8 and persons who are mentally or physically unable to indicate their quality of life, a caregiver can complete a proxy version of the EQ-5D.[22]

In some cases, the EQ-5D-5L may not be adequate. Alternative questionnaires and other methods for evaluating quality of life may therefore be supplied alongside the reference case, if it is argued that the EQ-5D-5L is insufficient in the situation in question. In this situation, generic outcome measures are preferable to disease-specific outcome measures. The alternatives are discussed in the **in-depth module 'QALY en kwaliteit-van-leven-metingen'** (only available in Dutch).

3.3.2 Disutilities

Certain interventions, such as those involving medicines or surgical procedures, can lead to side effects or complications which may impact the patient's quality of life. It is important to distinguish between side effects or complications with long-term consequences and short term consequences.

When measuring effects of complications and side effects on the quality of life - usually referred to as disutilities - the EQ-5D-5L (or EQ-5D-Y) questionnaire must be used. The questionnaire must be completed at the moment the side effects or complications occur. If this is impossible, the values (disutilities), as well as the corresponding durations, should be based on scientific literature.

3.3.3 Quality of life of informal caregivers

In some cases, interventions not only have an effect on patients, but also on their informal caregivers. Informal caregivers can fulfil an essential role in caring for a patient. Informal care is referred to as generally unpaid and often long-term care for sick family members or friends. This

may be care or help with daily activities. If the intervention affects the overall required number of hours of informal care and/or the quality of life of the informal caregiver, the quality of life of the informal caregiver should be considered in a scenario analysis. The maximum period of time which should be used in this instance is the patient's life expectancy.

Several instruments and methods are available for measuring and evaluating the quality of life of informal caregivers. However, sufficient evidence about the validity and sensitivity of the available instruments used in this specific context does not always exist. The National Health Care Institute recommends that the effects on the health-related quality of life of informal caregivers should primarily be identified using the EQ-5D-5L questionnaire during a clinical trial. If this is not possible, valuations from the literature should be used (measured with the EQ-5D questionnaire). When the EQ-5D questionnaire is used to assess the quality of life of both the patient and the informal caregiver(s), the outcomes can be combined. If the quality of life of informal caregivers is based on valuations from the literature, it can be included in the economic evaluation as a function of the patient's health status.[23]

3.4 Consulting experts

If relevant data are lacking, experts can be consulted in order to obtain data for an economic evaluation. According to the National Health Care Institute, as much published data as possible should be used and experts should only be consulted if no other evidence is available. As a result, consulting an expert is regarded as the weakest form of evidence. Elements of the evaluation that are based on expert consultations should thus be investigated in sensitivity analyses.

Because expert opinions frequently differ, an expert panel consists of at least 5 independent people affiliated with various centres. In the case of very rare diseases a lower number of panel members is allowed. The experts should have sufficient knowledge of the subject on which they are being consulted. For that reason experts can be clinical experts but also, for example, experts with practical experience.

Information from experts should be collected and reported in a structured fashion (see also **paragraph 5.1.4** and **tables B1-B3 in annex I**). A distinction is made between 'expert opinion' and 'expert elicitation'.

3.4.1 Expert opinion

Expert opinion involves obtaining qualitative information from experts.[24] Examples include determining the status of a new medical intervention, information about current clinical practice and the course of a disease. Initially, information should be gathered from each expert individually. This can be accomplished through a personal interview, or by sending out a questionnaire. The experts must provide an explanation of each answer they give. If the individual experts provided conflicting answers, efforts must be made to reach a consensus. This can be done by organising a meeting of the consulted experts, or by applying the Delphi method.[25, 26] Clear minutes must be supplied in the event of a meeting, showing the views of the individual experts.

3.4.2 Expert elicitation

Expert elicitation involves obtaining quantitative values from experts.[24] For example, they may provide an estimation of a certain input parameter of an economic evaluation, such as the time an intervention is used, or a distribution of various follow-up treatments. Various structured methods exist for expert elicitation, such as the Sheffield Elicitation Framework (SHELF) method, the Delphi method, and the Cooke method.[26, 27, 25, 28] The use of a structured method is obligatory. These methods start with an individual elicitation, by an interview or questionnaire. A fixed interval method or a variable interval method can be used to request parameter values.[29] In addition to point estimates, these methods also enable to generate a distribution around the parameter values. Various tools are available for requesting estimates from experts.[30, 27]

In order for experts to answer the questions properly, a number of sample questions must be used to prepare them for the use and application of distributions and probabilities.[31] Steps

must be taken, for each requested parameter, to verify whether the outcomes actually reflect the experts' thoughts.[31] The manual by Horscroft et al. describes various ways to do this.[32]

After collecting the estimates of the individual experts, the outcomes must be combined, or aggregated. This can be accomplished through mathematical aggregation or behavioural aggregation.[29] When it comes to mathematical aggregation, the individual estimations are combined using an algorithm. In the case of behavioural application, interaction takes place between the experts in order to generate an outcome through consensus. This can be accomplished through giving feedback on the answers provided by the other experts (the Delphi method), or by organising an expert panel (the SHELF method). Both methods require not only a point estimator to be aggregated, but also a distribution. These distributions must then be included in the deterministic sensitivity analyses and the probabilisticanalysis (see **paragraph 4.7**).

4 Methods

4.1 Possible study designs

An economic evaluation can be performed as part of an empirical study, or by using a simulation model. An empirical economic evaluation mainly involves the use of costs and effects at patient level from a controlled study (with data not being extrapolated over time). If there is no empirical study available which is in line with the PICOTS (see **paragraph 2.3**), it is necessary to use a model. When it comes to a model-based economic evaluation, the expected costs and effects of interventions are estimated using a simulation model.

Good reasons need to be provided for the chosen approach (empirical approach, model-based approach).

4.1.1 Empirical economic evaluations

When it comes to an empirical approach, it is important that the clinical study is organised pragmatically in order to make a valid estimation of the costs and effects of the interventions. Recommendations relating to the format of empirical economic evaluations can be found in the ISPOR guidelines.[33]

In the statistical analyses of the data, a number of complexities must be taken into account, such as a data imbalance, correlated costs and effects, and clustering of data.[34, 33] If regression techniques are used to determine average costs and effects, clear reasons must be given for the regression technique and the regression model. Data imbalance must be taken into account by, for example, using linear regression in combination with non-parametric bootstrapping or generalised linear models. To account for the correlation between costs and effects, seemingly unrelated regression can be used. Multilevel regression models or generalised estimating equations should be used in the case of clustered data. An explanation of methods to apply in the case of data imbalance, correlated costs and effects and clustered data is described in a scoping review.[34]

In addition to the above-mentioned complexities, there may also be, for example, differences in baseline characteristics (bias) between the randomised groups and missing data. **Paragraph 4.4** explains in more detail how to correct for various types of bias. **Paragraph 4.3** describes how to deal with missing data.

4.1.2 Model-based economic evaluations

When a model-based approach is considered, a choice has to be made for a certain type of model. This choice depends greatly on, among other things, the research question and the nature of the disease. Commonly used models are state-transition models (also known as Markov models), partitioned survival models, discrete event simulations and dynamic transmission models. Various documents/tools can be used to choose the right type of model.[35-37] The input parameters for the model are estimated on the basis of empirical data (for example a clinical trial) for which recommended statistical methods.[38]

4.2 Discounting

Data should be discounted when it is collected or modelled over a period of more than one year. For the reference case, future costs must be discounted with a constant discount rate of 3% and future effects with a constant discount rate of 1.5%.

The discount rate of 3% for future costs is based on the discount rate of 2.25% which applies to social cost benefit analysis (SCBA),[39] and an upward adjustment that takes current market developments into account. Based on the idea that the value of health grows over time and that this is not taken into account in another way, the discount rate for effects is lower than the discount rate for costs. The discount rate for effects is calculated as the discount rate for costs minus the growth rate of the consumption value of health.[40, 41] The annual growth percentage of the consumption value of health in the Netherlands is estimated at 0.6% to 2.9%.[42] A growth percentage of 1.5% is used, which is slightly below the average of the range referred to

and is also the upper limit of a previous German estimation.[43] Based on a discount rate of 3% for costs, the discount rate for effects is therefore 1.5% (3% minus 1.5%).

4.3 Missing data

Various factors can result in missing observations. The way in which missing data is dealt with forms an integral part of the statistical analysis element of an empirical analysis, as well as the analysis of empirical data to estimate model parameters.

It is important to take account of missing data during the data analysis because this could otherwise result in distorted estimations and excessive or reduced precision.[44, 45] In the literature a distinction is made between simple methods (such as complete cases, mean imputation and last value carried forward) and more advanced statistical methods (such as multiple imputation and expectation-maximisation algorithm).[44, 34]

First, it is important to check which missing data mechanisms play a role. When it comes to small proportions of missing data (<5% missing data) or 'Missing Completely At Random' (MCAR), simple methods, such as complete case analysis, usually suffice. However, the proportion of missing data is often bigger, also assuming that the data are 'Missing At Random' (MAR) applies. In these instances multiple imputation must be used to impute the missing data. Given that it is impossible to distinguish between MAR or 'Missing Not At Random' (MNAR), it is also recommended to perform scenario analyses in order to test the robustness of the MAR assumption. Various methods are available for this purpose, such as selection models, pattern-mixture models and reference-based multiple imputation.[46]

It may also be the case that no data are available for the estimation of a parameter value. In that case expert elicitation can be used (see **paragraph 3.4.2**).

4.4 Adjusting for bias

The effects found in a comparative study may be biased. Using a randomised controlled trial (RCT) reduces the likelihood of biased results. However, some types of bias can still occur within an RCT. In the case of a non-randomised controlled trial, the risk of bias is rather high. This paragraph contains details on the various analysis methods which can be used adjust for bias.

4.4.1 Adjusting for bias in randomised trials

Various types of bias can occur in RCTs. Selection bias and attrition bias are the key types of bias for which an adjustment must be made in the analysis are selection bias and attrition.[47] Selection bias may result in differences in baseline characteristics between the randomised groups. In this case an adjustment must be made using regression techniques. This reduces bias and increases the precision of the estimates.[34, 33] Attrition bias arises if dropped out patients in a trial differ from patients who continue the trial. Attrition bias can be reduced by using imputation (see **paragraph 4.3**). In addition, an intention-to-treat analysis is always required. If it is necessary to correct for bias, a scenario analysis must be added using the non-adjusted analyses.

4.4.2 Adjusting for bias in non-randomised trials

The risk of selection bias in non-randomised controlled trials is considerable. In order to adjust for this bias, a propensity score analysis or matching techniques need to be used. Another major challenge is confounding when analysing the data. Adjustments should be made for this using adequate statistical methods.[48] A distinction can be made between measured and unmeasured confounding.[48]

Matching or inverse probability weighting can be used to adjust for measured confounding, with the aim to make the various treatment groups comparable.[49] Confounding can also be adjusted using techniques such as regression adjustment, multivariate regression, or propensity score adjustment.[50]

In order to adjust for unmeasured confounding, methods can be used which are covered by instrumental variable analysis. In that case a number of different options are possible, such as

two-stage least squares regression and instrumental propensity scores.[50] An alternative approach is to simulate randomisation by adjusting for the influence of the effect of confounders with difference-in-differences and regression-discontinuity analyses. There are yet more methods which can be used to minimise the risk of bias in non-randomised controlled trial.[49, 50]

If correcting for bias is necessary, clear reasons must be provided for the choice of method. It is important that underlying assumptions of the methods are taken into account, clearly described and, if possible, tested. If there is corrected for bias, a scenario analysis must be added which shows the results when non-adjusted analyses are used.

4.5 Extrapolation techniques

If the time horizon of a clinical trial is shorter than the relevant time horizon as defined in the PICOTS, extrapolation techniques must be used (in a model-based economic evaluation). Longitudinal time-to-event data (Kaplan-Meier curves) are primarily extrapolated by assuming an underlying parametric distribution. An estimation of the course of the data after observation can be made by estimating the parameters of this distribution based on the observed data. When doing so, various assumptions can be made about the course of events over time. All standard parametric distributions must be evaluated and reported: exponential, Weibull, Gompertz, log-logistic and lognormal, gamma and generalised gamma. More flexible methods such as spline and cure models are also possible, but these are only allowed if the standard parametric distributions are insufficient.[51, 52] Any decision to use these methods must be properly substantiated.

The following steps need can be taken to determine which distribution is best for extrapolating the observed data:

1) Evaluation of visual fit: to what extent does the estimated curve follow the observed Kaplan-Meier curve. In this case, a sufficiently detailed scale must be used for the vertical and horizontal axes.
2) Evaluation of the proportional hazards assumption: the proportional hazards assumption must be assessed in order to determine whether dependent or independent parametric distributions should be fitted to the time-to-event data. To do this, log-cumulative hazards plots are used which support the choice for a proportional hazards model. It is preferential to fit dependent parametric distributions, unless it is demonstrated that the proportional hazards assumption does not hold. If dependent parametric distributions are fitted, a single parametric curve will be fitted for both intervention arms, with the intervention serving as a clarifying variable. In the case of independent parametric distributions, the values of the parametric curves for both intervention arms will be fitted independently. It is therefore possible that different parametric distributions are chosen for the intervention arms. In both cases, it is important to provide reasons for the chosen approach and implicit assumptions, also on the basis of a clinical rationale.
3) Evaluation of the statistical fit: the statistical fit must be evaluated using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).[53]

The above steps focus on the internal validity of the parametric distributions used, but do not give any indication about how suitable these distributions are after the follow-up duration of a clinical trial (clinical and external validation). For that reason, the clinical and external plausibility of the extrapolations is more important than the internal validity. The following step is therefore very important:

4) Clinical and external validation: other data sources such as similar clinical trials with longer follow-up durations and longer registers, can be used to assess the clinical plausibility of the extrapolated curves.[54] If major differences between the extrapolation and external data exist, it is important to evaluate whether the differences are due to a suboptimal extrapolation, or limitations in the external data. If no external data is available, validation by clinical experts will suffice (see **paragraph 4.8**). It is essential to perform a clinical validation of the type of distribution and the underlying distribution of events over time (increasing, decreasing, or constant hazards). In addition, it is important to perform a clinical evaluation of the distribution of the incremental effect over time (does the majority of the incremental effect occur during the follow-up

duration of a trial or after observation), see **paragraph 5.2.2**. If the extrapolation results in a higher survival rate than that of the general Dutch population, an adjustment should be made.

In the case of immature data (when the median survival rate is not reached) a substantial extrapolation of the data is performed.[51, 52] This implies increased uncertainty, in addition to the already existing uncertainty related to the extrapolated data. Conservative assumptions should be made in these situations. From the most suitable and clinically plausible curves, the most conservative curve (in other words the one that results in the smallest incremental difference in life expectancy between interventions) should be chosen in the base case analysis.

The uncertainty related to the estimated parameters must be included in the probabilistic analysis (see **paragraph 4.7.2.2**). Moreover, scenario analyses must be conducted to evaluate how the results change with other relevant distributions and shorter time horizons (see **paragraph 4.7.2.3**).

4.6 Subgroup analyses

Within an indicated patient population there might be heterogeneity. This could lead to differences in incremental costs and effects between subgroups. Various types of heterogeneity exist which are relevant to perform subgroup analyses.[55] Important types, that can currently be included in reimbursement decisions, are related to patient or disease characteristics which are linked to the effectiveness of an intervention or the baseline risk of events occurring. If these forms of heterogeneity based on clinical rationale exist, whereby subgroups can be discerned in clinical practice and whereby it is ethical and in accordance with the principle of equality to (not) reimburse a different intervention on the basis of heterogeneity, subgroup analyses must be performed. Empirical studies must have sufficient statistical power for these subgroup analyses. If the statistical power is ultimately found to be insufficient, this is not a valid reason to only determine the average effects in the total population. The effects found must then still be applied per subgroup in subgroup analyses. All other parameters, such as complication probabilities, quality of life values and uptake of care, should also be determined specifically for the subgroups.

The cost-effectiveness analysis should be performed per subgroup. The uncertainty relating to the estimates of the subgroups should also be included in the uncertainty analyses.

4.7 Uncertainty analyses

The outcomes of an economic evaluation are surrounded by uncertainty. First, uncertainty may be caused by method-related choices. This is referred to as methodological uncertainty. This type of uncertainty can largely be overcome by following this guideline and using the reference case. There are also other types of uncertainty. The following paragraphs examine the different types of uncertainty and how they can be identified in the context of empirical and model-based economic evaluations.

4.7.1 Uncertainty analyses in empirical economic evaluations

In addition to methodological uncertainty, statistical uncertainty and uncertainty in terms of fixed parameter values (for example cost price estimates) may also be a feature of empirical economic evaluations.

Valid statistical methods must be used to quantify the uncertainty around total and incremental costs, effects and the incremental cost-effectiveness ratio (ICER). The skewness of cost data and any correlation between costs and effects should be taken into account. Bootstrapping is the most common method for estimating uncertainty within the total and incremental costs and effects and the ICER.[56] The number of sampling draws used in the bootstrap should demonstrably lead to a stable result. According to the literature, there should be at least 5,000 replications.[58]

Uncertainty in terms of fixed parameter values must be demonstrated in deterministic uncertainty analyses. Parameter values should be varied within a range which is based on the

standard error. If no information is available about the standard error and the parameter has a major effect on the outcomes, expert elicitation must be used (see **paragraph 3.4.2**). If no information is available and the parameter only has a minor effect on the outcomes, it is sufficient to assume a range of +/-20% (considering the logical limits of the parameter in question).

4.7.2 Uncertainty analyses in model-based economic evaluations

Uncertainty is often a feature of model-based economic evaluations due to a lack of data on an important outcome parameter, a small sample size and data immaturity. TRUST is a tool that can be used to identify the various uncertainties.[59]

In the case of microsimulation models it is important to clarify which size of the simulated population is required for the results to be stable.[60] This can be accomplished visually, for example, by plotting the relevant outcome(s) against an increasing simulated population size, possibly with a 95% confidence interval, or by estimating the standard error or confidence interval of the relevant outcome(s) on the basis of a predefined (large) simulated population size. If parametric distributions are used to reflect variability, the uncertainty within the parameters of these distributions must also be included. In the case of microsimulation models it is recommended that this uncertainty is included by means of bootstrapping, because this allows the correlation between the parameters to be maintained based on the same empirical dataset.[61] It is important to make a clear distinction between variability and other uncertainty.

In order to identify the effect of uncertainty in model-based economic evaluations, a distinction is made between deterministic and threshold analyses, probabilistic analyses and scenario analyses. Value of information (VOI) analyses are used to estimate the consequences of uncertainty.

4.7.2.1 Deterministic uncertainty analyses

A deterministic uncertainty analysis is used to provide insight into the effect of parameter uncertainty on the results. The simplest form is the univariate analysis, in which the value of a single parameter is varied within a range (based on the standard error) while the other parameters are held constant. If no information on the standard error is available and the parameter has a major effect on the outcomes, expert elicitation must be used (see **paragraph 3.4.2**). If no information is available but the parameter only has a minor effect on the outcomes, a range of +/-20% can be applied (considering the logical limits of the parameter in question).
In a multivariate analysis two or more parameters are varied at the same time.

A threshold analysis can be used to determine which value a parameter must have in order to change a decision, or to identify the value for a parameter which leads to a predefined result.[38]

4.7.2.2 Probabilistic analysis

The aim of a probabilistic analysis is to estimate the impact of all parameter uncertainties together on the outcomes of the model. This is done by varying all input parameters that contain uncertainty simultaneously and by using a Monte Carlo simulation to create a large number of estimations of the incremental effects and costs.

The distributions used to estimate the uncertainty relating to the input parameter, must be reported and justified. The choice of distribution also depends on the logical limitations of the parameter in question.[62] The ranges used in the probabilistic analysis must be equal to the ranges in the deterministic analysis (see previous paragraph).

Possible correlations between parameters should also be considered. The correlation must be maintained in the case of, for example, parameters which are part of the parametric distributions fitted to the Kaplan-Meier curves (see **paragraph 4.5**). This can be done, for example, by using the Cholesky decomposition method.[62]

In the case of microsimulation models, care must be taken to avoid mixing variability/heterogeneity and uncertainty.[38, 63, 64]

The number of iterations used in the probabilistic analysis should demonstrably lead to a stable result. Methods for evaluating this are described in the article by Hatswell et al.[65] A properly functioning random number generator and a fixed random seed value are essential in order to reproduce model outcomes.[66] The ICER should be calculated by dividing the mean incremental costs by the mean incremental effects across all probabilistic analysis iterations.

More information about the probabilistic analysis can be found in the **in-depth module 'Onzekerheid en Value of Information analyses'** (only available in Dutch).

4.7.2.3 Scenario analyses

Scenario analyses should be used to identify the effect of methodological choices and key assumptions. The following scenarios must always be performed for both empirical and model-based economic evaluations:

i) no discounting;
ii) healthcare perspective (see **paragraph 3.2**), without the indirect costs of life years gained.

In the case of model-based economic analyses, the following scenarios must also be performed:

i) alternative lengths of the time horizon (at least: the duration of the clinical trial and halfway the extrapolated time horizon).
ii) other relevant distributions for extrapolation of time-to-event data (if this is part of a model-based economic evaluation).

The following scenarios should be performed in the described situations:

iii) other sources for costs and care uptake (if there are plausible alternative sources and the chosen values are uncertain).
iv) other sources for quality of life data (if there are plausible alternative sources and the chosen values are uncertain).
v) the non-inclusion of unrelated future medical costs (in the event of life-prolonging interventions);
vi) the inclusion of quality of life of informal carers (if this is influenced by the intervention, see **paragraph 3.3.3**).

Other relevant scenario analyses can also be added. All scenario analyses must be performed on the basis of a probabilistic approach so that the related results can be compared with those of the standard probabilistic analysis.

4.7.2.4 Value of information analysis

A Value of Information (VOI) analysis provides additional insight into the effect of uncertainty on decision-making.[67] A VOI analysis is based on the results of the probabilistic analysis (see **paragraph 4.7.2.2**). The analysis focuses on quantifying the consequences of uncertainty included in the analysis and therefore supports decision-making on whether or not to postpone the reimbursement of a new intervention and any additional research.

The expected value of perfect information (EVPI) and expected value of partial perfect information (EVPPI) should be determined as part of the reference case. The EVPI is the expected value of perfect information across all modelled aspects of the decision-making problem, which is equal to the expected costs of uncertainty associated with making the decision based on the current imperfect evidence.[67] The EVPI is also a risk criterion which shows the consequences of uncertainty. The EVPPI shows the effect of individual parameters or a group of parameters on the consequences of the total decision-making uncertainty. Various approximation methods exist for the determination of the EVP(P)I (see **the in-depth module 'Onzekerheid en Value of Information analyses'** (only available in Dutch)).

All parameters included in the probabilistic analysis should also be included in the EVPI. In the EVPPI, parameters can be combined into groups based on the uncertainty relating to this group of parameters. When choosing for groups of parameters it is important that there is some logical cohesion between the selected parameters, meaning that information on the parameters can be collected in the same study. Examples of groups of parameters are valuations for the quality of life, all cost parameters, the effect estimations and the extrapolations of time-to-event data.

The EVPI and EVPPI are multiplied by the size of the population that benefits from the study so that the population EVP(P)I can be calculated. It is therefore necessary to estimate the size of the population to which the decision relates. At least a time horizon of 5 years should be adopted. Furthermore, another time horizon can be included if supported by reasons, such as historical evidence and anticipated future changes (for example patent expiry, other studies and future new interventions for the same population).

If the population EVP(P)I is small (i.e. lower than the costs of additional research), it means that conducting additional research is most likely not the optimal option. Decision-making could in this case be based on the current information. In the event of a higher population EVP(P)I (indicating that decision-making uncertainty will have substantial consequences) it can be decided to perform additional VOI analyses. Additional VOI analyses (expected value of sample information (EVSI) and expected net benefit of sampling (ENBS)) can be used to determine whether, and what kind of, additional research is worthwhile and whether the definitive decision should be postponed while awaiting the results of additional studies. The need to perform these additional analyses depends on the extent of the consequences of the decision-making uncertainty and the feasibility of performing additional studies. These additional analyses are therefore not an obligatory element of the reference case. For more information, see the **in-depth module 'Onzekerheid en Value of Information analyses'** (only available in Dutch).

4.8 Validation

4.8.1 Validation in conjunction with a model-based economic evaluation

In the case of a model-based evaluation, it is important that various elements of the model are validated. The aim of the validation is to establish the usability and reliability of the results. The validation should encompass the conceptual model, the input data, the software-aided implementation (code verification/technical validation) and the model outcomes (operational validation). The ISPOR-SMDM task force refers to various types of validation tests.[68] This paragraph describes which validation tests must be performed for each element. The AdVISHE checklist must be used to report the validation.[69]

4.8.1.1 Validation of the conceptual model

To validate the conceptual model, at least cross validation must be performed.

Cross validation involves comparing the model with other models.[68] A literature review must be carried out into relevant studies in which similar indications and/or interventions have been modelled. Not only published scientific literature should be reviewed, but also earlier assessments by the National Health Care Institute (if available). Possible explanations for any identified differences should be addressed.

Face validation should be performed in case differences are found between the particular model the models in the studies and earlier assessments, or if a comparison with other models is not possible. The face validation involves an assessment from several (clinical) experts on model elements (for example the model structure). It is important that the presentation of the elements to be validated corresponds with the experts' knowledge and experience. A description of the face validation performed must be provided, detailing the views of the individual experts with regard to the elements submitted. The consulted experts should provide an insight into possible (conflicts of) interests, using the declaration of interest drawn up by the Royal Netherlands Academy of Arts and Sciences (KNAW).

4.8.1.2 *Validation of the input data*

Validation of the input data involves a thorough verification of the preprocessing of raw data used to estimate model parameters.[38, 69] In the case of statistical analyses (for example regression analyses) the usual tests must be carried out (analysis of discrepancies between statistical model and observations, a normality test where relevant, etc.).

4.8.1.3 *Technical validation*

The technical validation of a model (also referred to as (code) verification) involves checking the correctness of all model calculations. *Techver* is an instrument which can be used for this.[70]

4.8.1.4 *Operational validation*

Operational validation involves the validation of model outcomes. The outcomes that should be validated is determined on the basis of the clinical relevance and their effect on the cost-effectiveness results. Both internal and external validation are important in this. The internal validation compares the model results with input data. The external validation compares the model results with empirical data which has not been used as input data. A literature review of relevant publications is required for the external validation, similar to cross validation. If the results of the model differ from the literature findings, possible explanations for those differences must be examined in detail.

In addition, the model outcomes must be validated by means of cross validation. If differences are observed between the outcomes of the model and the outcomes of other models and empirical data, face validation is also required.

4.8.2 Validation of an empirical economic evaluation

The quality of empirical studies is guaranteed by regulations governing the performance of these studies (good clinical practice (GCP)). Nevertheless, in the case of empirical economic evaluations it is also important that the pre-processing of the data, assumptions relating to uptake of care, parameters used, the statistical analyses and the outcomes are validated as described in the previous paragraph. This indicates that cross validation and face validation are required for empirical economic evaluations as well, including checking whether correct parameter values (for example cost prices) have been used. Furthermore, whether the assumptions made in the empirical economic evaluation correspond with assumptions in previously published similar economic evaluations could be evaluated. The preprocessing of raw data, such as the data cleaning and the corresponding assumptions should also be validated (validation input data), as well as the statistical analyses underlying the outcomes (technical validation). Finally, the outcomes of an empirical economic evaluation should be compared with other empirical data (operational validation and external validation).

5 Reporting

5.1 Reporting data

5.1.1 Effectiveness

The results of the clinical trials which form the basis for the effectiveness estimates in a model-based economic evaluation must be reported. For all economic evaluations, the characteristics of the patient population studied (age, gender, baseline data of clinically relevant parameters, etc.), as well as the effectiveness of the intervention and the comparative intervention must be specified.

If estimates are based on systematic reviews, (network) meta-analyses and indirect comparisons, the methods used should be reported and justified in a transparent manner.

5.1.2 Costs

The prices and volumes of all costs components should be reported separately. The year to which the prices relate must be reported as well.

5.1.3 Quality of life

The following characteristics should be reported for each quality of life valuation: the instrument or technique that has been used, the respondents' nationality and the perspective (patient or societal). Moreover, an overview of the moments at which the quality of life has been measured should be provided. In empirical economic evaluations the quality of life must be reported for each of these moments in a table or graph, including the corresponding ranges.

In a model-based economic evaluation, the valuations for each health status must be presented separately. When these valuations are based on the literature, the following characteristics must be reported per valuation: the questionnaire or valuation technique used, the respondents' characteristics and the perspective (patient valuations or societal valuations).

Reasons must be provided for any deviation from the EQ-5D questionnaire (EQ-5D-5L of EQ-5D-Y with Dutch valuation[20, 21]). If quality of life data from the literature is used, it must be explained why this can be generalized to the target population. If quality of life outcomes of additional instruments are also presented, the validity of each instrument must be substantiated with scientific literature.

5.1.4 Consultations with experts

Requirements apply to the reporting of an expert opinion and expert elicitation. An overview of the requirements is shown in **tables B1-B3 in annex I**.[24] A KNAW declaration of interest from each consulted expert must be provided.

5.2 Reporting the methods

The report must clearly formulate and justify the PICOTS and the study design (empirical or model-based economic evaluation).

5.2.1 Missing data

Each economic evaluation based on an empirical study should include a description on whether there is any missing data and how this is taken into account in the analyses. It is important to report the amount of missing data and whether patients with missing data differ from patients without missing data (which missing data mechanisms apply, in other words censoring, MCAR, MAR or MNAR, see **paragraph 4.3**). If missing data has been imputed, the methods used should be reported.

5.2.2 Extrapolations

If time-to-event data has been extrapolated (see **paragraph 4.5**), at least the following must be reported, so the validity of the extrapolation can be assessed:

- The visual fit of the extrapolated curve with the observed data;
- An evaluation of the proportional hazards assumption;
- The statistical fit based on the AIC and BIC for each distribution studied;
- A clinical validation of the extrapolation;
- The percentage of patients who are alive at various points in time, using various distributions.

In addition, the percentage of the estimated effect that occurs during the follow-up period must be reported. This is calculated by dividing the incremental QALYs over the follow-up period of the study (the time period for which time-to-event data is available) and the incremental QALYs for the entire time horizon of the model:

$$Ratio = \frac{\Delta QALY\ for\ the\ follow-up\ duration}{\Delta QALY\ for\ the\ entire\ time\ horizon\ of\ the\ model}$$

The same calculation must be made to report the percentage of the estimated after patients stop using the intervention (if applicable). This is calculated by dividing the incremental QALYs during the treatment period by the total incremental QALYs for the entire time horizon of the model.

5.2.3 Validation

The AdVISHE checklist must be completed and submitted in order to report the validation of a model-based economic evaluation.[69]
In the case of empirical economic evaluations at least the following elements must be reported:

- Validation of input data (element B1 of AdVISHE)
- Validation of outcomes (elements D1, D2 and D3 of AdVISHE)

For the statistical validation the results of the usual tests must be reported (analysis of discrepancies between statistical models and observations, a normality test where relevant, etc.).

The experts consulted for face validation should have completed the KNAW declaration of interest.

5.3 Reporting the results

5.3.1 Base case analysis

The reported results should contain the absolute effects and costs, as well as the incremental effects and costs, for both the intervention and the comparative intervention(s). Confidence intervals must be reported for these outcomes as well. The ICER (expressed in costs per QALY gained) and the net monetary benefit (NMB) must be reported. The results should be based on the probabilistic analysis. The absolute and incremental costs should be broken down by cost category (see **table 3** in **paragraph 3.2.1**) and presented as total sums. The absolute and incremental health effects must also be broken down according to health status in the case of model-based economic evaluations.

In terms of effects, QALYs must be presented in any case, regardless of the length of the time horizon. In the case of a long time horizon in which survival plays a role, life years must be reported as well.

A reference value (cost-effectiveness threshold) is required for the calculation of the NMB. This can be based on the burden of disease which applies to the indication concerned. See in this respect the report entitled **'Kosteneffectiviteit in de praktijk**' (only available in Dutch).[71]

When comparing more than two interventions, a fully incremental analysis should be carried out for the reporting of the outcomes. In a fully incremental analysis, the results of the various interventions are ranked in a table from low to high costs or QALYs. Following that, an indication of which interventions are strongly and weakly dominated is provided.[72] The ICERs can be calculated for the interventions which are not dominated. The results can be presented using an efficiency frontier.

The same must also be done if the optimal use of a specific intervention or strategy is being investigated, such as the optimal screening interval or the optimal cut-off value of a diagnostic test. It is important to evaluate a sufficiently large number of values for the parameter(s) to be optimised.[73]

Ranking based on the NMB can also be used to determine the most cost-effective intervention.

5.3.2 Subgroup analyses

All model outcomes in **paragraph 5.3.1** must be presented separately for each subgroup.

5.3.3 Deterministic and scenario analyses

The effects of variation in the input parameters in the deterministic uncertainty analyses should be presented in a well-structured way. The reporting of the deterministic analyses for parameter uncertainty should include the lower and upper ICER (discounted) in conjunction with the chosen distribution of the parameter, as well as the incremental costs and effects. The chosen ranges for the distribution must be clearly specified and explained. It is recommended that the results are presented in a table and that the most influential parameters are presented graphically in the form of a tornado diagram. Results are always presented together with, and compared with, results of the base case analysis.

The results of scenario analyses should be presented in a table, listing the incremental costs and effects and the ICERs.

5.3.4 Visualisation of probabilistic analysis

In a model-based economic evaluation or a (bootstrap) analysis of the statistical uncertainty in an empirical economic evaluation, results of the probabilistic analysis must be presented graphically in a cost-effectiveness plane (CE-plane). An indication of the percentage of the (bootstrap) iterations that fall in each of the four quadrants must be reported.

The results of the probabilistic analysis must also be presented in a cost-effectiveness acceptability curve (CEAC).

5.3.5 Value of information

The results of the VOI analyses should be presented in an EVP(P)I curve. The EVP(P)I must at least be reported for the population. The size of the population to which the decision relates, as well as the relevant life expectancy of the intervention, must be reported and substantiated.

The EVP(P)I curve must contain a variety of reference values. If the ICER is higher than the applicable reference value (cost-effectiveness threshold),[71] the range of values on the x-axis must contain the value of the ICER.

If EVSI and ENBS calculations have been conducted, the results of various study designs and for the various reference values must be presented. For more information please refer to the **in-depth module 'Onzekerheid en Value of Information analyses'** (only available in Dutch). More information about reporting a VOI analysis can be found in the CHEERS-Value of Information Reporting Standards.[74]

6 Areas of concern

Economic evaluations for decision-making in the Netherlands must be performed in accordance with the reference case (see **Chapter 1**). Specific areas of concern apply to a number of situations. These areas of concern are discussed in this chapter.

6.1 Sequential elements

The use of a certain intervention can affect the choice of a subsequent action. For example, in diagnostics and screening, different test results can lead to different types of follow-up examinations and other treatment strategies. In addition, using a certain intervention can affect the effectiveness of a subsequent intervention. Sequential consequences like these must therefore be part of the economic evaluation.

For model-based economic evaluations of diagnostics, where numerous successive tests are used, it is important to use conditional probabilities of test outcomes instead of independent probabilities.[75] Not only can a test lead to a (further) selection of tests (a negative test will not be followed by further tests, in contrast to a positive test), the probability of an outcome of a follow-up test is often dependent on the outcomes of one or more previous tests (for example the probability of a positive CT scan if lung cancer is suspected is, for example, greater after a positive x-ray than after a negative x-ray).

6.2 Incremental innovation

For some healthcare interventions, incremental innovation throughout their life cycle is expected. This is not exceptional in the case of medical devices, but it can also apply to other interventions. The intervention could be improved further during or after an economic evaluation has been performed. As a result it may be necessary to create a dynamic model for the effectiveness and the costs. It is important that the specifications of the care intervention investigated and any assumptions about (phased) innovation are accurately reported. The effect of any assumptions should be investigated in scenario analyses.

6.3 Learning curve

For many care interventions, such as medical devices, screening and surgical interventions, there is a learning curve which has an effect on the outcomes. This may have an impact on the external validity of the study results, such as when the findings of a study involving experienced users are generalised to the entire population. The opposite applies in the event that outcomes of a short-term study with inexperienced users are extrapolated over a longer time horizon. It is important that the effect of any learning curve is clarified using a scenario and/or uncertainty analysis, where the learning period is excluded.

6.4 Alternative outcomes

The primary aim of interventions in healthcare is not always to improve the health-related quality of life or extend the life expectancy of the patient. For some interventions, such as diagnostic tests and medical devices, effects may be achieved in terms of broader value components.[76, 77] Examples include convenience for the caregiver, or a reduction in the time required for the procedure. It may be relevant for these kinds of interventions to include those other value components in the economic evaluation as well. Methods are available which can be used to identify and quantify that valuation by directly consulting patients and users, such as a discrete choice experiment (DCE) and the Multi-Criteria Decision Analysis (MCDA).[78-82]

In the context of long-term care, the primary aim of the intervention to be evaluated is not always health-related quality of life, as measured with the EQ-5D-5L and/or to prolong lives, but well-being instead. As a result, it may be appropriate to use well-being as an outcome parameter in these cases. Examples of instruments to measure well-being are the ASCOT, ICECAP-A, WiX, WOOP or the EQ-HWB.[83-88] In addition to the EQ-5D, a validated well-being questionnaire with a descriptive system which generates a utility value (on the basis of Dutch tariffs) is advised to use in a scenario analysis.

For informal carers it also applies that the effects of interventions may be broader than health alone. Instruments measuring broader outcome measures among informal carers include, for example, the CES, ASCOT-carer or CarerQoL.[89-91] These can be presented in scenario analyses.

Many interventions in youth care or mental healthcare are aimed at improving behaviour, or reducing psychological problems. Such interventions can have an impact on the patient's mental health as well as those around them or their family circumstances. As a consequence, the effects of interventions frequently extend beyond the patient's health. Improving the functioning of the family, or lowering the level of public nuisance or drug use, for example, can be important goals that necessitate extra attention to the careful measurement of the effects. Broader outcome measures (as described above) or outcome measures aimed at mental health (which can be measured, for example, using the MHQoL or ASC-TASI),[92, 93] may be suitable in this respect. More information can be found in the 'Manual of intersectoral costs and benefits'[16] and in the 'guide to costs-benefits analysis in the social domain'.[18] It is important to avoid double counting.

7 References

1. Zorginstituut Nederland. Beoordeling stand van de wetenschap en praktijk 2023. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2023/04/11/beoordeling-swp-2023.
2. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 1: introduction. Med Decis Making 2013; 33: 597-606.
3. Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013; 33: 607-17.
4. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making 2013; 33: 618-40.
5. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making 2013; 33: 641-56.
6. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making 2013; 33: 657-70.
7. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Med Decis Making 2013; 33: 671-8.
8. Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338: b1147.
9. Jansen JP, Fleurence R, Devine B, et al. 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 Health 2011; 14: 417-28.
10. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14: 429-37.
11. Chaimani A, Caldwell D, Li T, et al. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from: www.training.cochrane.org/handbook.
12. EUnetHTA. Methods guideline - D4.3.2 Direct and indirect comparisons. 2022. Available from: via https://www.eunethta.eu/wp-content/uploads/2022/08/EUnetHTA-21-Deliverable-D4.3.2-Methodological-Guideline-on-Direct-and-indirect-comparisons-V1.0.pdf.
13. Drummond M, Sculpher M, Claxton K, et al. Methods for the economic evaluation of health care programmes. Fourth edition. Oxford: Oxford University Press, 2015.
14. Kellerborg K, Perry-Duxbury M, de Vries L, et al. Practical Guidance for Including Future Costs in Economic Evaluations in The Netherlands: Introducing and Applying PAID 3.0. Value Health 2020; 23: 1453-61.
15. Hoefman R, Van Exel N and Brouwer W. iMTA Valuation of Informal Care Questionnaire (iVICQ). Version 1.1 (December 2011). Rotterdam: iBMG / iMTA, 2011. Available from: https://www.imta.nl/questionnaires/ivicq/documents/.
16. Drost R, Paulus A, Ruwaard D, et al. Handleiding intersectorale kosten en baten van (preventieve) interventies: Classificatie, identificatie en kostprijzen. Maastricht: Maastricht University, Department of Health Services Research, 2014.
17. Pokhilenko I, Janssen LMM, Paulus ATG, et al. Development of an Instrument for the Assessment of Health-Related Multi-sectoral Resource Use in Europe: The PECUNIA RUM. Appl Health Econ Health Policy 2023; 21: 155-66.
18. Koopmans C, Heyma A, Hof B, et al. Werkwijzer voor kosten-batenanalyse in het sociale domein. Amsterdam: SEO Economisch Onderzoek, 2016.
19. Bouwmans C, Krol M, Severens H, et al. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. Value Health 2015; 18: 753-8.
20. Versteegh M, Vermeulen K, Evers S, et al. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health 2016; 19: 343-52.

21. Roudijk B, Sajjad A, Essers B, et al. A Value Set for the EQ-5D-Y-3L in the Netherlands. Pharmacoeconomics 2022; 40: 193-203.
22. EuroQoL. EQ-5D-Y - Available modes of administration. 2023. Cited 2023 Aug 24. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/.
23. Pennington BM. Inclusion of Carer Health-Related Quality of Life in National Institute for Health and Care Excellence Appraisals. Value Health 2020; 23: 1349-57.
24. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. Pharmacoeconomics 2016; 34: 1161-72.
25. Dalkey N and Helmer O. An Experimental Application of the Delphi Method to the Use of Experts. Management Science 1963; 9: 458-67.
26. European Food Safety Authority (EFSA). Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014; 12: 3734.
27. Oakley JE and O'Hagan A (2019). SHELF: The Sheffield Elicitation Framework (version 4.0). Cited 2023 Sep 11. Available from: https://shelf.sites.sheffield.ac.uk/home.
28. Cooke R, Mendel M and Thijs W. Calibration and information in expert resolution; a classical approach. Automatica 1988; 24: 87-93.
29. Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technol Assess 2021; 25: 1-124.
30. University of York - Centre for Health Economics. Structured expert elicitation resources (STEER). Cited 2023 Sep 11. Available from: https://www.york.ac.uk/che/research/teehta/elicitation/steer/.
31. Bojke L, Soares MO, Claxton K, et al. Reference Case Methods for Expert Elicitation in Health Care Decision Making. Med Decis Making 2022; 42: 182-93.
32. Horscroft J, Lee D, Jankovic D, et al. Structured expert elicitation for healthcare decision making: A practical guide. Date unknown. Available from: https://www.york.ac.uk/che/research/teehta/elicitation/steer/
33. Ramsey SD, Willke RJ, Glick H, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health 2015; 18: 161-72.
34. El Alili M, van Dongen JM, Esser JL, et al. A scoping review of statistical methods for trial-based economic evaluations: The current state of play. Health Econ 2022; 31: 2680-99.
35. Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value Health 2012; 15: 796-803.
36. Jin H, Robinson S, Shang W, et al. Overview and Use of Tools for Selecting Modelling Techniques in Health Economic Studies. Pharmacoeconomics 2021; 39: 757-70.
37. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Med Decis Making 2012; 32: 678-89.
38. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Making 2012; 32: 722-32.
39. Werkgroep discontovoet. Rapport Werkgroep discontovoet 2020. Available from: https://open.overheid.nl/documenten/ronl-d985cdc5-5da7-4644-be09-4c22bfa885be/pdf.
40. Gravelle H and Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Econ 2001; 10: 587-99.
41. Brouwer WB, Niessen LW, Postma MJ, et al. Need for differential discounting of costs and health effects in cost effectiveness analyses. BMJ 2005; 331: 446-8.
42. de Boer P, van Exel J, Fransen M, et al. Gezondheidseconomische aspecten van de Covidpandemie. De Bilt, RIVM rapport 2023-0318, 2023. Available from: https://www.rivm.nl/publicaties/gezondheidseconomische-aspecten-van-covidpandemie
43. John J, Koerber F and Schad M. Differential discounting in the economic evaluation of healthcare programs. Cost Eff Resour Alloc 2019; 17: 29.
44. Mutubuki EN, El Alili M, Bosmans JE, et al. The statistical approach in trial-based economic evaluations matters: get your statistics together! BMC Health Serv Res 2021; 21: 475.
45. Burton A, Billingham L and Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clinical Trials 2007; 4: 154-61.
46. Leurent B, Gomes M, Faria R, et al. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. Pharmacoeconomics 2018; 36: 889-901.

47. Phillips MR, Kaiser P, Thabane L, et al. Risk of bias: why measure it, and how? Eye (Lond) 2022; 36: 346-8.
48. Zorginstituut Nederland. Deliverable 3 Regie op Registers voor Dure Geneesmiddelen: Kader voor analysemethoden (D3). 2020. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2020/06/24/deliverable-3-regie-op-registers-voor-dure-geneesmiddelen
49. Faria R, Hernandez Alava M, Manca A, et al. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available from: http://www.nicedsu.org.uk.
50. Varga AN, Guevara Morel AE, Lokkerbol J, et al. Dealing with confounding in observational studies: A scoping review of methods evaluated in simulation studies with single-point exposure. Stat Med 2023; 42: 487-516.
51. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from: http://www.nicedsu.org.uk.
52. Rutherford M, Lambert P, Sweeting M, et al. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. Available from: http://www.nicedsu.org.uk.
53. Collett D. Modelling survival data in medical research. New York: Chapman and Hall/CRC, 2015.
54. Royston P and Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013; 13: 33.
55. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics 2008; 26: 799-806.
56. Glick H, Doshi J, Sonnad S, et al. Economic Evaluation in clinical trials. Second edition.Oxford University Press, 2014.
57. Efron B and Tibshirani R. An introduction to the bootstrap. New York: CRC Press, 1994. Geraadpleegd op via.
58. Andrews D and Buchinsky M. On the number of bootstrap repetitions for BCa confidence intervals. Econometric Theory 2002; 18: 962-84.
59. Grimm SE, Pouwels X, Ramaekers BLT, et al. Development and Validation of the TRansparent Uncertainty ASsessmenT (TRUST) Tool for Assessing Uncertainties in Health Economic Decision Models. Pharmacoeconomics 2020; 38: 205-16.
60. Hahn G. Sample Sizes for Monte Carlo Simulation. IEEE Transactions on Systems, Man, and Cybernetics 1972; 2.
61. Degeling K, MJ IJ, Koopman M, et al. Accounting for parameter uncertainty in the definition of parametric distributions used to describe individual patient variation in health economic models. BMC Med Res Methodol 2017; 17: 170.
62. Briggs A, Claxton K and Sculpher M. Decision modelling for health economic evaluation.Oxford University Press, 2006.
63. Vemer P, Goossens LM and Rutten-van Molken MP. Not simply more of the same: distinguishing between patient heterogeneity and parameter uncertainty. Med Decis Making 2014; 34: 1048-58.
64. O'Hagan A, Stevenson M and Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. Health Econ 2007; 16: 1009-23.
65. Hatswell AJ, Bullement A, Briggs A, et al. Probabilistic Sensitivity Analysis in Cost-Effectiveness Models: Determining Model Convergence in Cohort Models. Pharmacoeconomics 2018; 36: 1421-6.
66. Gentle J. Generation of Random Numbers. In: Härdle W, editor. Computational Statistics – Statistics and Computing. New York: Springer, 2009.
67. Fenwick E, Steuten L, Knies S, et al. Value of Information Analysis for Research Decisions-An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value Health 2020; 23: 139-50.
68. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making 2012; 32: 733-43.

69. Vemer P, Corro Ramos I, van Voorn GA, et al. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. Pharmacoeconomics 2016; 34: 349-61.
70. Buyukkaramikli NC, Rutten-van Molken M, Severens JL, et al. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. Pharmacoeconomics 2019; 37: 1391-408.
71. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk. 2015. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk
72. Paulden M. Calculating and Interpreting ICERs and Net Benefit. Pharmacoeconomics 2020; 38: 785-807.
73. O'Mahony JF, Naber SK, Normand C, et al. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. Value Health 2015; 18: 1138-51.
74. Kunst N, Siu A, Drummond M, et al. CHEERS Value of Information (CHEERS-VOI) Reporting Standards - Explanation and Elaboration. Value Health 2023.
75. Novielli N, Cooper NJ and Sutton AJ. Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? Value Health 2013; 16: 536-41.
76. Wurcel V, Cicchetti A, Garrison L, et al. The Value of Diagnostic Information in Personalised Healthcare: A Comprehensive Concept to Facilitate Bringing This Technology into Healthcare Systems. Public Health Genomics 2019; 22: 8-15.
77. Lee DW, Neumann PJ and Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. Value Health 2010; 13: 310-4.
78. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health 2016; 19: 1-13.
79. Marsh K, M IJ, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health 2016; 19: 125-37.
80. Lancsar E, Fiebig DG and Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. Pharmacoeconomics 2017; 35: 697-716.
81. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 2014; 32: 883-902.
82. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics 2019; 37: 201-26.
83. van Leeuwen KM, Bosmans JE, Jansen AP, et al. Dutch translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT). Health Qual Life Outcomes 2015; 13: 56.
84. Rohrbach PJ, Dingemans AE, Essers BA, et al. The ICECAP-A instrument for capabilities: assessment of construct validity and test-retest reliability in a general Dutch population. Qual Life Res 2022; 31: 687-96.
85. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. Value Health 2022; 25: 482-91.
86. Voormolen DC, Bom JAM, de Bekker-Grob EW, et al. Development and content validation of the 10-item Well-being instrument (WiX) for use in economic evaluation studies. EsCHER Working Paper Series No. 2023004, Erasmus University Rotterdam. 2023. Available from: https://www.eur.nl/en/research/escher/research/working-papers.
87. Bom JAM, Voormolen DC, Brouwer WBF, et al. Construct validity, reliability and responsiveness of the 10-item Well-being instrument (WiX) for use in economic evaluation studies. EsCHER Working Paper Series No. 2023005, Erasmus University Rotterdam. 2023. Available from: https://www.eur.nl/en/research/escher/research/working-papers.
88. Hackert MQN, van Exel J and Brouwer WBF. Well-being of Older People (WOOP): Quantitative validation of a new outcome measure for use in economic evaluations. Soc Sci Med 2020; 259: 113109.
89. Rand SE, Malley JN, Netten AP, et al. Factor structure and construct validity of the Adult Social Care Outcomes Toolkit for Carers (ASCOT-Carer). Qual Life Res 2015; 24: 2601-14.

90. Al-Janabi H, Flynn TN and Coast J. Estimation of a preference-based carer experience scale. Med Decis Making 2011; 31: 458-68.
91. Brouwer WB, van Exel NJ, van Gorp B, et al. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. Qual Life Res 2006; 15: 1005-21.
92. van Krugten FCW, Busschbach JJV, Versteegh MM, et al. The Mental Health Quality of Life Questionnaire (MHQoL): development and first psychometric evaluation of a new measure to assess quality of life in people with mental health problems. Qual Life Res 2022; 31: 633-43.
93. Reckers-Droog V, Goorden M, Kaminer Y, et al. Presentation and validation of the Abbreviated Self Completion Teen-Addiction Severity Index (ASC T-ASI): A preference-based measure for use in health-economic evaluations. PLoS One 2020; 15: e0238858.

Annex I

Table B1: Reporting criteria for an expert opinion

Criterion	Description
Research rationale	The reasons why a consultation with an expert is needed must be described. A systematic literature review must show that the evidence is not available in the literature.
Research question	A clear research question must be defined.
Data collection	There must be a clear description of how the questions have been put to the experts (for example by telephone, face-to-face, or by means of a questionnaire).
Questions	A copy of the questions asked must be submitted.
Composition of expert panel	The composition of an expert panel must be described. The description must clearly indicate which area of expertise each member represents and the centre they are affiliated to.
(Conflicting) interests	Possible conflicting interests must be described for each expert using the KNAW declaration of interest.
Meeting	If a meeting has taken place, clear minutes must be submitted, with a clear description of any differing views held by the participants.
Results	The outcomes must be clearly described. The answers provided by the individual experts must be reported (including a substantiation per question), as well as the (lack of) consensus achieved during any meeting of experts.
Interpretation	The interpretation of the (lack of) consensus achieved must be described.

This table is partly based on the article by Iglesias et al.[24] If the Delphi method is used for an expert opinion, the criteria from the article by Iglesias et al.[24] must be used instead of the criteria in this table.

Table B2: Additional reporting criteria for an expert opinion in the event of a Delphi study

Criterion	Description
Data collection	There must also be a description of how feedback on the results of previous rounds has been provided to the group and/or individuals based on the answers provided.
Rounds	The number of scheduled rounds and the number of rounds that have taken place should be reported, in combination with a description of the plan for proceeding to the next round.
Results	The results of each round must also be described.

These criteria apply in addition to the criteria shown in **table 1**.

Table B3: Reporting criteria for an expert elicitation

Criterion	Description
Research rationale	The reasons why a consultation with an expert is needed must be described.
Research question	All input parameters for which experts are to be consulted must be described.

Type of parameter	The type of input parameter (for example ratios, probabilities, distributions etc.) must be described.
Composition of expert panel	The composition of an expert panel must be described. The description must clearly indicate which area of expertise each member represents and the centre they are affiliated to.
(Conflicting) interests	Possible conflicting interests must be described for each expert using the KNAW declaration of interest.
Elicitation method	A description must be provided of how the data is to be collected (for example by an interview or a questionnaire).
Clarification/training	The use of clarifying and training material for the experts must be reported and made available.
Questions/assignments	The actual questions asked of the experts must be reported.
Feedback	It must be reported how, for each question, it has been verified whether the answer actually reflects the thoughts of the expert and/or the expert has revised his/answers following this feedback.
Aggregation method	The aggregation method used must be described.
Performance of experts	The weighting of the outcomes of the various experts and how this has been applied must be described. Differential weighting (on the basis of seed variables) or equal weighting can be used.
Results	The individual and aggregated estimation(s) of points and distribution(s) for each parameter must be presented.
Interpretation of the results	The interpretation of the outcomes and how these are used in the economic evaluation must be described.

This table is based on the article by Iglesias et al.[24] For more information about the various criteria, reference is made to the article in question.

Colophon

Contact person
H.A. Geuzinge
ageuzinge@zinl.nl

Address
National Health Care Institute
PO Box 320
1110 AH Diemen, The Netherlands