Technical Document

Modelling

Version 1.0

09.10.2008

Contact:

Institute for Quality and Efficiency in Health Care (IQWiG)
Dillenburger Straße 27
D-51105 Cologne
Tel.: +49 (0)221 / 35685-0
Fax: +49 (0)221 / 35685-1
E-mail: knb-methoden@iqwig.de
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Handling Uncertainty and Variability</td>
<td>40</td>
</tr>
<tr>
<td>11.1</td>
<td>Types</td>
<td>40</td>
</tr>
<tr>
<td>11.2</td>
<td>Quantifying Uncertainty</td>
<td>41</td>
</tr>
<tr>
<td>11.3</td>
<td>Recommendation</td>
<td>42</td>
</tr>
<tr>
<td>12</td>
<td>Validation</td>
<td>44</td>
</tr>
<tr>
<td>13</td>
<td>Model Documentation</td>
<td>46</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>48</td>
</tr>
</tbody>
</table>
1 Introduction

Estimates of the net economic effects of health technologies, including their potential impact on budgets, are essential. In most cases, economic data collected in clinical trials will not be sufficient on its own to ensure fully informed decision making. Cost data from clinical trials are not very useful in most cases because the study environment so modifies practice that applicability suffers. Moreover, they are unlikely to be specific, or adequately transferable, to Germany, and in any case, will not provide information on the longer term consequences of adopting the new technology [1,2]. Indeed, economic data from clinical trials may not be available at all. Thus, modelling of the economic outcomes is an essential component of the evaluations.

This technical document describes modelling approaches in health care and provides guidance on the process of developing economic models, modelling techniques, model validation, uncertainty analysis, and reporting of modelling studies.

Although IQWiG may in principle be asked to assess different types of health technologies, such as preventive, diagnostic and therapeutic procedures, this technical document mainly focuses on the technical issues of economic modelling of therapeutic procedures.
2 Definition

Several definitions have been offered for the term “model” as applied in the context of health care. Models are analytic tools used to understand real-world systems, estimate outcomes for a given set of inputs and examine the effects of changes to the system being modelled. In effect, any evaluation that extends beyond direct application of observed data can be considered a model [3], and even direct application usually involves some form of statistical modelling. It is understood that models cannot represent reality perfectly: they are based on a reduced set of components and require simplifying assumptions [4]. Nevertheless, it is crucial that the model be valid in the sense that it sufficiently reflects the system it represents.

ISPOR (the International Society for Pharmacoeconomics and Outcomes Research) defines modelling as “an analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources” [5]. A recent description of modelling techniques adopts a similar definition: “a formal quantified comparison of health technologies synthesizing sources of evidence on costs and benefits” [6]. Both definitions are compatible with the requirements of economic evaluations conducted by IQWiG.
3 Purpose

Several situations have been described in which modelling should be used in the economic evaluation of health care [7]. The ultimate purpose of modelling the economic effects of a health care technology is to provide a firm basis for decisions on the reasonableness of reimbursing that technology at a particular price. Within the context of decision making in Germany, the patient-relevant benefits of treatment are based on health outcomes as analyzed by IQWiG, using internationally recognized standards of evidence-based medicine. Patient-relevant benefits of diagnostic and therapeutic interventions are identified in controlled clinical trials. In some cases valid models can predict how these benefits identified in clinical trials appear under other circumstances or time perspectives. However, under no circumstances can models generate new benefits. Thus, in Germany the purpose of modelling is mainly focused on the cost consequences of the health care technology in various scenarios. Only if the benefit assessment, which takes place before the economic assessment, already requires modelling, then modelling techniques will be used to estimate both health outcomes and costs. This may be the case if the benefits of the compared technologies have multiple attributes that must be combined in a score integrating the health consequences (such as quality-adjusted life years or other combined measures) (see Section 2.3.1 in the Main Document) or the benefit must be transferred to a cardinal scale that reflects how valuable that benefit is. This transfer may involve modelling to address for example if a longer time horizon has to be considered (see Section 2.3.1 in the Main Document).

This information is used to position the technology with reference to the efficiency frontier (see Section 3) and to estimate budget impact. This does not mean that health outcomes are not included in IQWiG’s economic models, but rather that modelling is not primarily used to estimate health outcomes per se. However, health outcomes must be part of the modelling, because without this component it is not possible to produce meaningful and reliable estimates of cost, and economic evaluation would be reduced to a deficient comparison of acquisition costs amongst competing therapeutic options.

Critics of the use of modelling in evaluations have argued that these analytic structures are subject to bias in selecting inputs, structural design and assumptions and can be impenetrable — a so-called black-box — which only the developers of the model can understand. While a given model may display all of these negative attributes, it is possible to produce a clear
framework with explicit assumptions and the best possible inputs given the available data [8]. Modelling also permits exploration of various health care utilization scenarios and strategies [4]. To be of use, however, it is essential that modelling studies undertaken by IQWiG to be:

1. fully transparent, with model inputs and assumptions defined and justified;
2. of sufficient depth to adequately represent the disease being modelled and the costs associated with it and the health care treatments at issue;
3. flexible enough to assess multiple scenarios under varying sets of assumptions and settings;
4. allow for an assessment of uncertainty in predicted costs;
5. and be populated with data that are relevant to Germany, including not only costs, but also clinical practice patterns, demographics and epidemiology.

Models must also undergo rigorous validation, both in the sense of the integrity of internal calculations, and in terms of external validity. Therefore IQWiG models should undergo review, whereby the reviewers are provided access to both the relevant technical documents and to a fully functional and evaluable electronic version of the model.
4 Process

The process to follow in model development, validation, analysis and reporting is outlined in the remainder of this technical document. Briefly, the recommended steps involved in model development are as follows:

1. Define the research question(s) that the model will be expected to address. This includes:
   a. What therapeutic options will be assessed?
   b. What population or populations are relevant to the analysis?
   c. What are the relevant cost categories associated with the therapeutic area and the interventions evaluated (see Section 4)?

2. Develop an influence diagram which graphically describes the relationships between key components of the problem, and how they interact with each other. The influence diagram is meant to provide an overview of the disease and its economic aspects and to summarize the critical relational components of the planned model.

3. Closely related to the development of the influence diagram is specification of the model concept in terms of the flow diagrams that serve as the “blueprint” for the model. The flow diagrams should elucidate how the health effects of the evaluated therapeutic options are linked to concrete health outcomes and resource use, and how they are projected over the modelling time horizon. The concept should explain how the relationships in the influence diagram will be dealt with, and provide justification for these choices. The concept will influence the modelling technique used to answer the research question (and not the reverse – i.e., commitment should not be made to a modelling technique first).

4. Data to populate the model should be selected only after the concept has been established. It is not good practice to design the model around available data, as this may be inordinately restrictive and will tend to produce models of insufficient depth, and leading to lack of attention to data collection and other drawbacks. Thus, although data availability may eventually restrict implementation of the model concept, it should drive neither the concept itself nor the influence diagram unless the data directly contradict these. Rather, the model should reflect the concept and influence
diagram that were felt to best capture the problem at issue and the full functionality should be built into the model. The analyses can then be run under restricted data conditions. This allows for more flexible exploration of outcomes under hypothetical sets of relations and the potential impact these might have.

5. If the data are insufficient to populate the model that has been conceptualized, targeted data collection should be considered. This may take the form of specific observational studies or may leverage existing registries or even databases that accumulate information for other purposes (see Section 8.2.4).

6. Assessment, and possibly collection, of data is followed by development of the functional relationships that form the core of the model. The type of relationships to be quantified will be driven by data availability and modelling technique selected, but it is essential that these mathematical relationships be fully documented. In the case of *de novo* equations (e.g., derived from statistical regression modelling), documentation of their development is required.

7. The modelling technique that will be used to structure the model should be chosen concurrently with development of the functional relationships, as these activities are closely interrelated. The main criteria for selection of technique should be its ability to encode the model concept and influence the diagram as designed, and thus answer the relevant research questions by adequately representing the system being modelled. While transparency is also essential, even very complex models can be transparent with adequate documentation. Oversimplification in models, however, is fatal. No matter how transparent the construct, if the model does not adequately represent the system, it will not be useful [9].

8. Implementation and programming of the model should only proceed upon validation of the model concept, selection of data, development of functional relations to be used and choice of modelling technique.

9. During implementation, technical validation should already be in progress. This is completed, along with face validation, upon finalization of the programming.
10. After validation of the model is completed, analysis of the relevant therapeutic options and target populations can proceed. These analyses must cover the most likely scenarios and also address areas of uncertainty in the inputs or structure [10].

11. Reporting should encompass all of the process items described above. There can be considerable leeway in descriptions of the modelling methods, as long as they are clear and complete. Forcing a fixed template may compromise the research team’s ability to describe the model in the fullest and clearest way. Reports must be accompanied by a fully functional and accessible electronic model. The model must be made publicly available in order to ensure confidence in its integrity.

Developing a valid model is a substantial undertaking, and as such, the use of previously validated models is encouraged. IQWiG will retain a bank of models developed under its purview, which can be used for further economic evaluations. Nevertheless, modifications and updating of existing models will be required to maintain currency.
5 Influence Diagrams

5.1 Definition

Influence diagrams are being increasingly used for the development of economic models [11].

An influence diagram is like a drawing that reflects the key concepts for the model. It contains the important features of the disease being modelled and shows the relationships among them. Despite its name it is not per se a causal diagram; rather it reveals the associations between aspects that must be considered. This is also not per se Bayesian, although it does convey the state of knowledge at the time of model design.

5.2 Purpose

An influence diagram is created to describe and communicate the key relationships at the core of the model and to display the most important parameters which will drive the model. By making these explicit and visual, the diagram allows all concerned to understand what the modellers envision and to recommend changes without having to comprehend the technical aspects of the model itself. This ensures that everyone is clear about what will be modelled and, if changes are made later (in response to data constraints, for example), about what has to be given up and how much distortion in the relationships will be introduced.

The influence diagram also provides the basis for developing detailed flow diagrams of the model to inform the development process.

5.3 Components

The influence diagram must include all the important features of the disease being modelled and the relationships amongst these including:

- Patient characteristics
- Pathophysiology and clinical processes of the disease
- Treatment and disease management
- Outcomes

These components are typically depicted using oval shapes, and the relationships between them using arrow connectors with the arrowhead indicating the direction of the relationship
(without necessarily indicating causality). The diagram may become quite complex, but comprehensiveness is the objective, not simplicity. Colour and other devices may help clarify the relationships depicted.

In later stages of modelling, the influence diagram can also help summarize data sources at-a-glance by labelling connectors with the data key sources (e.g., clinical trial, literature, database analysis) as these are selected.

An explanation of the components and connections in the influence diagram should be provided in a written document accompanying the diagram. In this document, each component (i.e., oval shape) and connection is explained. The rationale for omitting connections that might have been included should also be given.
6 Model Concept

6.1 Definition

The model concept provides the detailed specification of the analytic structure and approach. It follows from the influence diagram and provides the intended design in much greater depth.

6.2 Purpose

Defining the model concept is a critical step in model development, both in terms of transparency and for determining validity. It must be clearly documented and presented. Its purpose is to ensure that the disease area being modelled and the effects of the interventions have been accurately understood and that any assumptions about relationships between model components are explicit. As even the most sophisticated models are simplifications of reality, with required assumptions and limitations on what can be included [10,12], only by clearly specifying the concept can the model’s validity and utility be properly understood.

6.3 Components

Specifying the model concept should take into consideration the following elements:

- What clinical events are to be included and what influences their occurrence?
  - Can these events recur, and what influences the risk of recurrence?
  - Does experiencing an event influence the likelihood or timing of subsequent events?
  - Does the severity of events needs to be evaluated?

- Which conditions are to be considered?
  - Do the conditions sufficiently describe the clinical spectrum of the disease?
  - Are they sufficient to discriminate patients regarding health status, health-related quality of life, and costs?
  - Do they sufficiently capture the patient’s history and experience?
• How do the intervention’s benefits, as pre-identified by IQWiG, influence the disease course?
   o Are events postponed or avoided altogether and/or their severity reduced?
   o What are the underlying mechanism and assumptions which determine how the intervention affects the natural history of the disease?
   o What assumptions on long-term effectiveness of the interventions, and on consequences of treatment delays or discontinuation?

• How does the behaviour of providers or patients influence outcomes?
   o Is compliance, either by medical professionals or patients, important?
   o How are future medical decisions on patient management influenced by response to and/or compliance with treatment?
   o Is interaction between patients relevant, as it would be, for example, in the case of acute infectious diseases?

• How are costs influenced by patient management, patient history, patient behaviour, and the timing and severity of complications?

6.4 Validation

Models should be “consistent both with a coherent theory of the health condition being modelled and with available evidence regarding causal linkages between variables” [5]. Thus, finalization of the core model concept cannot proceed without validation.

In order to provide for validation of the model concept, it is important to include full documentation of the information used to define the model concept and the hypothesized linkages. It is also helpful to compare the concept with that of relevant existing models (i.e., cross-model validation). This comparison should include a description of, and justification for, any deviations from concepts normally applied in the disease and therapeutic areas in question.
Concordance with published economic evaluations, however, is insufficient given the uneven quality of published modelling studies and the lack of agreement on what constitutes a “good model” [13]. Models must make sense to people with knowledge of the disease [14]. Validation of the model concept, therefore, must include presentation of the full concept to appropriate experts, who should include individuals familiar with the disease areas being modelled, but not be restricted to clinicians. Relevant expertise would be epidemiologic, statistical, quality-of-life, public health and economics.
7 Mandatory Features

To be considered appropriate for IQWiG purposes, the costs estimated using models must conform to a minimum set of requirements. This section outlines those requirements, but it is noted that the quality of models cannot be dictated based on features alone, making peer-review and fully accessible models a critical feature of all such cost estimates to be considered by IQWiG.

7.1 Appropriate Depth

As detailed in Section 9, IQWiG’s models must have a sufficient level of depth to answer questions regarding the costs related to a particular health care technology. This will require

- capturing all aspects of the disease and treatments that are likely to have a meaningful impact on costs;
- including both benefits and harms of interventions;
- incorporating the relevant heterogeneity in the target population, natural history of the disease, and intervention effects;
- reflecting medical practice patterns in Germany;
- incorporating the effects of less than perfect compliance with treatment;
- allowing outcomes to be reported over time and by category of costs rather than simply as an aggregate cost over a fixed time interval; and
- allowing analyses that can incorporate different demographic, epidemiologic, cost and practice pattern data in order to assess outcomes that are meaningful to different regions of Germany.

Where sufficient data are not available to incorporate these features, it is essential to explore the possible impact on cost estimates of reasonable ranges of values for the missing information.
7.2 Perspective

The perspective of the analysis should be that of the community of German citizens insured by the SHI. This is not identical to the perspective of the SHI itself, as it includes resources of patients and society.

If possible, cost outcomes should be reported, both in aggregated and fully disaggregated, in at least the following categories (please note: these are not mutually exclusive):

- The health care technologies being evaluated (e.g., drug costs)
- Other activities directly attributable to the health care technologies being evaluated (such as monitoring and treatment of adverse events)
- Hospitalization
- Emergency department
- Outpatient physician visits
- Other professionals
- Equipment or device
- Other (see Technical Documents – Cost estimation, 2.2)

7.3 Time Horizon

The time horizon of the model should allow for the capture of all costs affected by the health care technologies at issue. This can be quite long, for example, lifelong in chronic diseases. Depending on the evaluated intervention, the time horizon can even extend beyond the lifetime of the people directly affected, for example, in the evaluation of vaccination. Thus, a reasonable duration must be chosen that covers the most important costs for the reimbursement decision. If a shorter time horizon is adopted, then the rationale for this decision must be explained.
7.4 **Duration of Treatment**

Treatment duration should be evaluated considering at least two scenarios:

- treatment lasts as long as it did in the trials from which treatment benefits were assessed, and

- the treatment duration involves the relevant time span for the reimbursement decision (e.g., the practice patterns in Germany).

Treatment duration must also take into consideration premature treatment discontinuation, whether due to patient behaviour, side-effects, medical advice, or lack of sufficient beneficial effect.

7.5 **Discounting**

All costs evaluated in the model must be reported at their appropriately discounted value. The reader is referred to Section 3.2.5.2 in the main document and to the technical document “cost estimation” on discounting costs. The sensitivity analysis should cover a broad range of discount rates to investigate the effect of discounting on the intervention’s net cost. Discount rates will be set by IQWiG, and updated if necessary.
8 Assessment of Data

The available data should be systematically researched, and their quality and relevance for supporting the model inputs and risk functions needs to be assessed. The quality criteria developed by epidemiologists for each level of evidence are generally considered by experts in this field to be equally applicable to assessing the data available to support economic studies [15], but this is incorrect. Thus, data from well-conducted randomized controlled trials are suggested as preferable to those from non-randomized studies. However, their relevance for cost information tends to be weak, if they do not correspond to usual practice [15]. Controlled prospective studies might also be preferred to retrospective ones, but for cost information a database cohort created retrospectively or a case control approach may also be useful.

Typically, data will be drawn from several sources, including trials for relative treatment effects, cohort studies for parameters and risk factors relating to the natural history of the disease and the associated life expectancy, cross-sectional surveys for quality of life data, resource use and costs, registries, automated data bases, or compiled statistics. Assumptions based on expert opinion should be avoided, as they are rarely accurate enough to be used when other sources are not available. An important criterion for the studies is the transferability of results to the German context.

The rationale behind any selections of specific data sources should be included (such as selection criteria used, external validity), especially for the key parameters which may have a substantial impact on the model results. Details of the methods of data processing to develop the required inputs, functions or distributions of parameters must be provided [13,16,17].
8.1 Objectives

To identify the data available, review its quality, assess its relevance for the specified model objectives and identify any data gaps for which assumptions will be appropriate; and apply a suitable method to process the data in order to develop reliable estimates.

8.2 Potential Sources

8.2.1 Trials

The clinical outcomes assessed by IQWiG are mortality, morbidity, health-related quality of life and validated surrogates. These clinical outcomes are provided given by the preceding IQWiG benefit assessment. To be judged acceptable per IQWiG criteria, any surrogates must be supported by studies that have demonstrated convincingly that changes in these clinical measures translate to changes in patient-relevant outcomes. Criteria to comply with for valid surrogates are described in IQWiG’s General Methods. It is an issue of the preceding benefit assessment to identify valid surrogates and if available to evaluate the effect of the regarded intervention on these surrogates. These surrogates are then an option for the economic evaluation via modelling. Patient-relevant benefits of diagnostic and therapeutic interventions are identified in controlled clinical trials. In some cases valid models can predict how these benefits identified in clinical trials appear under other circumstances or time perspectives. However, under no circumstances can models generate new benefits.

The model may use inputs that are based on the relative efficacy from one trial, on pooled trial datasets, or on meta-analysis of reported results, depending on the quality, relevance and availability of the data. Generally, at the time when new products are reviewed for reimbursement, only relatively small and short-term efficacy trials have been completed. The durations of these efficacy trials is often insufficient to detect significant differences regarding cost information. This tends to be a major limitation for economic evaluations because the latter need to reflect practices in health care — which may extend to broader time horizons [18-20].

Data from the trials should be extracted after confirming their relevance, specifically considering choice of comparators, outcomes measured, surrogate versus final outcomes, length of follow-up, countries participating in the study, exclusion criteria leading to under-
representation for sub-groups of patients, protocol-driven resource use, patient or physician compliance [18-20]. The steps taken to transform the data extracted to a form useful for incorporation into the model should be described and any adjustments or calibrations made transparent.

The clinical outcomes measured in a trial will be linked in the model to resource use, such as hospital admissions, procedures, and other services. In many cases it will be necessary to supplement the trial data with information pertinent to the German health care system, using data collected from non-randomized studies. The usual pattern of care delivered in each country reflects local reimbursement rules and availability of certain services, such as access to nursing homes and payments to family members acting as caregivers, and these will impact the location of care. Many trials are international studies, and the patterns of resource use, locations of care, as well as epidemiologic parameters influencing costs are known to vary considerably between countries and thus cannot be assumed without justication to be directly transferable to Germany [21,22]. Trial protocols may require resources to be provided that would not be provided in usual care, and so other data sources may be useful to identify resources commonly used in actual practice, as it is not appropriate to include protocol-driven costs in the cost estimates. The relevance to the current German health care system of any resource use details collected in a trial should be evaluated and specifically addressed in the submission, as this may greatly impact the final conclusions of the study [23].

Meta-analyses of trials are often used to establish the relative treatment effects, and details of the methods used in the meta-analysis should be reported. Head-to-head comparisons of each treatment are not always available, so that indirect comparisons could be necessary to ensure all the pertinent treatments are considered. Appropriate methods should be used to derive these indirect estimates of treatment efficacy [24-26].

A transparent and complete documentation of the data analyses or data processing undertaken to develop the required inputs is considered to be a quality ensuring measure and a quality criterion for modelling practice.

8.2.2 Epidemiologic Studies

Data from observational studies are frequently necessary to augment those collected during a trial or to extrapolate from the trial results and support modelling beyond the trial time.
horizon. For many diseases it takes years or decades until the final health outcomes become manifest. In such cases it is common to assess intermediate clinical endpoints that serve as surrogates for the final endpoint that really matters to the patient such as long-term morbidity and mortality [7]. These surrogates must be supported by studies that have demonstrated convincingly that changes in these intermediate clinical endpoints translate to changes in final patient-relevant outcomes.

The interpretation of any prediction function, and the validity of transferring established risk functions to the modelled population, as well as any adjustments made, have to be explained and justified. Rational for the sources of any observational data should be provided. Details must be provided of the statistical methods applied and of any analyses undertaken to develop estimates in a form appropriate for incorporation into the model. The validity of assuming that these functions or estimates can be applied to the population considered in the model for Germany needs to be addressed and any modifications detailed.

8.2.3 Automatically Generated Databases

Automated claims databases can provide details of health care resource use by a large number of patients in actual practice. They are thus a useful data source when the diagnostic codes employed by the database are sufficient to specify the relevant patient population. This may be problematic if the definition includes clinical criteria, laboratory results, patient characteristics beyond age and gender or other items that are not typically encoded in such databases. Data on claims for medications can also provide information on real-world treatment paths, switching, compliance (at least with obtaining prescriptions if not with actually using the medications), and persistence.

The relevance of a given database to a specific model has to be assessed, as many factors can impact prescribing practices, including local formulary requirements, product launch dates, and even the process of releasing data, as there may be a delay until information is available for analysis. Although a given database may be useful for quantifying resource use patterns for a given condition, great caution must be exercised in comparative analyses, as there may be strong confounding by indication when patients are not randomly assigned to treatments [16].
8.2.4 Registries

Patient registries are observational studies that systematically collect a reduced set of data from large numbers of patients managed in the routine care setting. As the data are usually collected in the regular care setting, they will tend to reflect actual clinical practice more accurately and provide a good source of information. Unlike trials, registries tend not to mandate admissibility criteria or treatment options to the same extent.

A well-designed registry can be a valuable source of data for a model. It can support analyses to develop many of the required predictive functions describing the association between clinical events and resource utilization, quality of care, work days missed, and so on.

Assessments of the value of data from a registry should include evaluating its relevance, given the registry objectives, design, patient population(s), and conduct, in order to confirm the quality and completeness of the data collected (for example, by checking for the proportion of missing data on important variables and understanding how these were handled and reported) [27]. The analyses undertaken to develop functions and estimates that are in a form appropriate for inclusion into the model must be detailed. Additionally, appropriate access to the registry must be provided in terms of quality control.

8.2.5 Compiled Statistics

Compiled statistics are health statistics compiled by governments from various sources such as a census or survey. These sources might provide relevant data on population demographics such as age, sex, and health-relevant behaviour, weight, and so on [28]. Compiled statistics are often used to estimate mortality for causes other than those specifically of interest in the model (e.g., the Human Mortality Database [29]). Any sources of compiled data used in the model must be provided and assessed in terms of quality and relevance to ensure that these are appropriate, reasonably current, and complete.

8.2.6 Expert Opinion

Expert opinion, focus groups and consensus panels can be used in a very limited way in modelling studies: to inform model structure and support assumptions made in the influence diagrams and model concept. It has been said that expert opinion may be used also in situations where there are constraints on the data available and research resources to conduct
additional studies to obtain more precise estimates [13,30]. In these circumstances, it is said that a pragmatic approach may be necessary, and the evidence available can be supplemented at that time with informed opinion [13,17,30]. Others have stated that conducting economic analyses using expert opinion will be considered by assessors as a methodological flaw [31].

For actual input values, and even more so for functional relations, this source of information is almost never appropriate [16,32-34]. It is recommended that modelling studies carried out by IQWiG minimize the use of expert opinion and insist on obtaining actual data (if possible, during the assessment) to support estimates.

Expert opinion may be considered valuable regarding what is done (not how frequently) in the management of patients and may facilitate transferring data collected on patterns of resource use between countries when equivalent services are not available. The justification for use of expert opinion and a description of the methods used to elicit it must be included in the model report [32] and reflect any variability in the opinions obtained [13,32,34,35].
9  Functional Relations

9.1  Definition

A functional relation is an equation relating a response variable to its determinants, that is, to variables that are correlated or associated with the response. The covariates will typically include patient characteristics and time. These functional relations are derived from the data sources used to support the model employing appropriate statistical techniques. Alternatively, published equations may be used if they are suitable for the population modelled. This section describes the types of relationships that should be considered and statistical strategies suitable to derive these.

9.2  Purpose

Functional relations should be specified to reflect the relationships posited in the influence diagram of the model. That is, these equations should be used to ensure that parameters in the economic model that were felt to be associated are appropriately linked, so that a change in one parameter is reflected by a change in the other (here termed effect without necessarily implying causality). Ignoring relationships between parameters, or equivalently, omitting important factors from the functional relations specified in the model, leads to an averaging of the effect or response variable over levels or values of the omitted factors. Incorporating correlates of effect and costs parameters leads to more precise estimates, thus reducing the uncertainty of the economic results [36,37].

9.3  General Concepts

Functional relations are particularly important in simulation-based economic evaluations as they control the simulation. Multiple regression techniques and other methods can be used to create these functional relations [36,38,39].

Functional relationships should be specified to relate effect and cost parameters to patients’ characteristics, treatment, time and other factors that may be pertinent. At its simplest, the effect parameters would be related to the single determinants of interest only: thus, when addressing the intervention, these would yield “average” responses for the intervention and for whatever alternative was considered. Including patient characteristics that are associated
with the effect, on the other hand, allows patient-specific predictions of changes in the affected parameter. This framework also allows the integration of subgroup effects through the inclusion of interaction terms. Resource use can be modelled similarly to capture variations among patient types; for instance, older patients or those with more severe illness may incur greater costs. The statistical properties of resource use data present serious challenges and need to be handled carefully [40].

When several effects tend to be correlated, functional relations should be specified to link these outcomes. For example, the effect of a patient characteristic, such as the availability of a caregiver, may be considered in terms of the hospitalization rate and of the use of care services at home; large reductions in hospitalization are likely to be accompanied by large increases in home care services and vice versa. To link these, response on one outcome may be included as a covariate in the functional relation for the other endpoint, for example.

In economic models where patients are simulated by assigning characteristics from distributions observed in the data source, correlations between characteristics also need to be considered. For instance, weight is associated with the patients’ sex; risk factors tend to correlate, severity of illness may be predictable from patient characteristics, and so on. Capturing correlations between all characteristics may be difficult; instead, bootstrapping patient profiles from the data source may be more reliable, as given a suitably large and varied sample, this ensures all relationships are simultaneously captured.

In situations where effects are measured longitudinally or in terms of a time-to-event, functional relations must reflect how the measure per se changes over time; in addition, it is important to consider whether the effects vary over time. In the context of survival (or time-to-event) analyses, one of the most commonly used response measure is the hazard of the endpoint (death, or some other event); the patterns of changes in the hazard over time are determined by the type of statistical distribution assigned to the event times. In longitudinal analyses that address aspects other than occurrence of events (e.g., change in glycaemia), changes in the response over time should be explored using a variety of functional forms of the time parameter. In both cases, the effects of another factor need to be evaluated at different points in time; changes in the effect should be assessed and incorporated in the functional relations between the endpoint and covariates. When the data were collected from multiple locations or sites, it is also important to examine the role of location on effect and
cost parameters [41,42]. Confounding must be controlled for using an appropriate technique [42-45].

Continuous outcome variables such as blood pressure, weight, etc., can be modelled with linear regression techniques as long as their distribution is reasonably approximated by the normal. This is not the case, for instance, for data on costs, which tend to have a highly skewed distribution. Log transformations or alternative models based on gamma distributions, should be considered [38,42,46]. When the skewness in costs is caused by an excess number of patients with no costs, a two-step approach in which the probability of incurring a cost is modelled first, followed by analysis of the magnitude of the costs among patients incurring some expense [47].

In some cases, economic models require a prediction of the timing of an event, to estimate duration of use of a treatment, for instance. Non-parametric or parametric analyses, e.g., failure-time analysis, are required in these situations to describe how the hazard of the event changes over time. This is done by specifying a statistical distribution for the event times; the most commonly used distributions are the exponential, the Weibull and the Gompertz [48]. The exponential distribution assumes the event occurs at a constant rate or hazard and thus tends not to be useful for medical applications given the fact that most hazards in medicine are not constant over time. The Weibull distribution assumes the hazards have a monotonically increasing or decreasing shape over time (with the exponential as a special case). The hazards in a Gompertz distribution also increase or decrease monotonically but at a much faster rate than in the Weibull distribution. The log-normal or log-logistic distributions should also be considered [48]. All of these distributions, except the exponential, involve a \textit{shape} and a \textit{scale} parameter. Either parameter can be allowed to vary according to patient characteristics and other determinants.

Evidence supporting the validity of the functional relations used in the economic model should be presented. This should include description of all the assumptions and decisions made in developing the equations. Sufficient information should also be provided about the data sources to show that these were adequate to support the types of analyses that were conducted.
Various strategies may be used to select variables when developing a prediction equation. These include automated selection procedures (forward, backward, stepwise selection), or manual selection of variables based on statistical significance, or prior clinical knowledge about the relationships being examined. In accordance with its methods (Allgemeine Methoden 3.0, [49]) IQWiG recommends backward selection. As complete absence of clinical knowledge about the relationships is extremely rare, it is recommended that prior clinical knowledge is taken into account in any model building procedure. In any case, the approach used to select the variables for each equation should be described and justified.

Explicit information should also be included about the criteria used in selecting the form of the equations. For instance, in the case of a linear equation the underlying assumption can be that the criteria may be to maximize explained variability. In logistic and failure-time analyses, predictive functions may be evaluated in terms of their discriminating power; measures such as the c-index or receiver-operating characteristic (ROC) curves can be used [50].

The assumptions invoked in building the functional relations should be presented, along with approaches used to ensure that the assumptions were verified. The goodness-of-fit of chosen regression models should be investigated and described by means of common procedures [50,51]. When developing causal functions, the assumptions on the existence and directions of causal links should be explicitly stated using causal graphs [52].
10 Choice of Technique

A number of modelling techniques are available and have been applied in health economic evaluations such as decision trees, Markov state-transition models, discrete event simulation, transmission models, agent-based simulation and others [3,6,12,53-57]. It is extremely important that a commitment to a particular technique not made in advance. Hence, IQWiG has no a priori preference for a specific modelling technique. The choice of the appropriate modelling technique depends on the research questions posed by the Federal Joint Committee (G-BA) through commission to IQWiG. The properties of the evaluated technology, disease and setting are further criteria for the choice of modelling technique. The proper process to yield efficient models that address the estimation problems at issue is to fully understand and design the model and assess the available data and functional relations first. Only then should a modelling technique be chosen.

The ultimate choice of technique will be influenced by a number of factors, but the guiding principle is that the economic model should be adequate to answer the research questions being posed. Simplicity of design for the sake of minimizing process and analysis time, or to render the model understandable to non-experts, is unacceptable if this makes the model less able to inform real decisions. Equally unacceptable is choice of a modelling technique just because of familiarity with it. With adequate documentation, even “complex” models can be fully transparent, and increased computer processing capabilities and variance-reduction techniques mean that time and computing requirements to run complex models are considerably less demanding [9,57-59]. When the choice of technique leads to modification of the model concept, a red flag should be raised — the choice is very likely to be the wrong one.

The modelling techniques described in the following sections are by no means meant to be exhaustive, but are the most commonly used in health economics. Readers are referred to Brennan’s taxonomy of model structures [6], Koopmans examination of infectious disease models [60], and Stahl’s overview and guide on modelling methods [56].
10.1 Aspects of the Choice

Although many characteristics of models could be considered in defining the choices to be made [6], two that have been prominent in health economic evaluations are the level at which to model the population (the entire group or the individuals that conform it) and the basic concept used to represent the system (based on the states people can be in or on the event they can experience).

10.1.1 Cohort versus Individual Models

Broadly speaking, models can be divided into those that address individual patients and those that do not distinguish individuals, using instead a cohort approach. Cohort models aggregate the individuals into a group which becomes the unit of analysis. The initial group is defined by a single set of characteristics (e.g., age, gender, presence of disease). As things happen over time this group “breaks up” into pre-defined subgroups according to the experiences being modelled. These subsequent subgroups are also defined based on their own set of characteristics (e.g., new level of disease). Individual models consider the experience of each patient separately, even if they report results at the level of the population. Each individual has a unique set of characteristics and is modelled over the time horizon with his or her events and experiences recorded. Although these individual models typically require more data, computing power, and running times [57,58], this does not necessarily mean that they are more difficult to specify and understand.

Cohort models are the most common technique used in health economics today [12,61]. Individual simulation (often termed “microsimulation”), once nearly unheard of in health economic evaluations, has gained considerably in prominence in recent years [57,58]. Although there continues to be resistance to its use [62], it is widely accepted that in many situations, individual-level simulation can provide more flexibility and accuracy and microsimulation models can often be used for a wider range of health care questions. On the other hand, if the individual patient characteristics and their heterogeneity in the target population are not of great relevance to the decision problem, cohort simulation may be sufficient to answer the research question [3,6,9,55,57,58,63-65].
For IQWiG health economic evaluations, the choice of cohort simulation versus individual-level simulation must be carefully justified. If a cohort technique is employed, it must be demonstrated that the choice of a cohort approach did not inappropriately modify the model concept and that it does not introduce distortions that represent significant departures from reality. If an individual-level simulation is employed, it must be demonstrated that appropriate distributions and covariance matrices have been used for the individual model parameters and that the assumptions on these distributions are not unrealistic simplifications of reality.

10.1.2 Event-based versus State-transition-based Simulation

Broadly speaking, two approaches have been used to conceptualizing the health economic system under study. The most commonly used technique in modelling in general is one that considers the events that can happen — discrete event simulation [66] and related techniques [67]. In health economic evaluations, however, the state-transition concept (Markov modelling) is far more prevalent today [61]. State-transition models have been introduced in health care evaluation in the early 1980’s [68] and have a long history of use in health care decision making, including clinical and economic applications [69].

In a Markov model, the population moves through several (mutually exclusive) health states in specified time intervals (Markov cycles). During this process, events, health outcomes (e.g., years of life, QALYs) and costs are accumulated and can be compared between the interventions of interest [70]. In a state-transition model, either the entire cohort starts at time zero in an initial disease state or the cohort can be distributed between different Markov states. At each cycle of the model, the appropriate transition probabilities are applied and the distribution of patients in each state of the model is recalculated. Running this analysis for many cycles builds up a profile of how many patients are in each state of the model over time.

State-transition models require several simplifying assumptions and limitations. Patients (whether as fractions of a cohort or as individuals) are described in terms of the conditions they can be in [69]. These discrete “states” are mutually exclusive and assumed to be homogeneous (i.e., everyone in a given state is identical). Therefore, a sufficient number of states must be chosen to represent the full spectrum of the disease and if a person acquires two or more properties (e.g., history of myocardial infarction and stroke), Markov states must be created for such combination states. Moreover, time moves in discrete jumps, so it is
important to choose a sufficiently small cycle length that allows for subsequent moves between Markov states in a short time if this is realistic in nature.

A strong assumption of Markov models is that there is no memory (the “Markov Assumption”). In other words, persons who move from one state to another acquire all the characteristics of the receiving state and lose all prior ones. The likelihood of moving from one state to another depends solely on the current health state. In particular, it can depend neither on prior experience nor on duration of time in a state. This means that relevant experience such as prior events must be included in the description of the states. For example, so-called “tunnel states” must be created to reflect the time a patient has already spend in a specific disease state. This can easily lead to an explosion of the number of states and make the Markov model unwieldy. In these cases, it may be useful to use individual-level simulation with first order Monte Carlo simulation because this allows for “tracking” the history of each individual during the simulation, and hence overcoming the restrictive (and often unrealistic) Markovian assumption. In such an implementation, further transition probabilities, costs and other parameters can be modelled as a function of the entire patient history [69].

Event-based simulation [66] conceptualizes reality in terms of the events that can occur. This turns out to be a very flexible concept that allows for natural representation of the disease process, treatment patterns, compliance with treatments, and other relevant factors. As time is explicit and continuous and events can be anything that happens in reality and can occur in whatever order makes sense, with no a priori restrictions, the experience of patients can be captured very accurately. Thus, events can reflect changes in clinical conditions, medical procedures and visits, the use of other resources, patient behaviour, decision points, and even changes to the system as a whole (e.g., the introduction of a new treatment, or a change in reimbursement). In addition, these simulations allow individuals not only to possess their own characteristics but to “carry” their personal histories. This permits the model to keep track of what has happened and use this to alter what may happen next. Thus, time since onset of some condition, duration of treatment, response to treatment, prior events and decisions, and any other relevant past occurrences can affect the subsequent course, as appropriate. As resources and their characteristics — including capacity — are modelled explicitly, these
simulations provide a very realistic platform for estimating costs. It can be stated, that event-based models are powerful tools that are currently underused in health economics.

Given the flexibility and realism of event-based simulations, it is recommended that these should be considered at least in the same way as state-based models for IQWiG evaluations. For example, if within a day, an individual might experience a stroke event, an admission to hospital event, and a death event, and this should be captured by the model, an event-based approach may be the best choice, as it allows for more accurate application of costs and discounting [58].

The rationale for the choice between the state-based and the event-based modelling approach must be detailed and it must be shown that the choice of a particular technique did not inappropriately modify the model concept or introduce unacceptable distortions of reality.

10.2 Techniques

10.2.1 Decision Trees

A decision tree is a visual representation of all the investigated technology options and the consequences that may follow each one [71]. Each intervention is followed by branches representing the possible consequences with their respective probabilities. Probabilities may depend not only on the intervention but also on patient characteristics (e.g., subgroups with different risk factor profiles). At the end of the tree each path leads to an outcome, for example, the endpoint of a clinical trial, a combined score for symptoms and side effects, quality of life, disease-free survival or overall survival. For each intervention, the expected value of the clinical and economic consequences can be calculated as a weighted average of all possible consequences, applying the path probabilities as weights. Decision trees work well for very simple decision problems, analyzing events with limited recursion and a limited and fixed time horizon. Decision trees can be evaluated either with cohort simulation or with patient-level simulation.

10.2.2 Discrete Event Simulation

In the area of health economics, the use of discrete event simulation has become more prominent in recent years [56]. Nevertheless, existing books [72] and manuals [73,74] deal
with examples from many other fields but do not venture into health economics beyond consideration of the physical organization of health care services [75-79]. Nevertheless, the basic approach to discrete event simulation is well covered in the non-medical literature [80]. In this section, a brief description of the key elements of this technique is provided, along with the application of these to health economic evaluations.

Discrete event simulation operates at the level of individual *entities* which are the objects that will experience the events considered in the model. There can be many kinds of such objects in a discrete event simulation, and the entities may even change type during the simulation. In health economic models, a nearly universal type of entity is the patient, but the basic entity could also be defined at the level of organs in a much deeper model. Other common entities are health professionals, caregivers and even inanimate objects. Entities are created and assigned a type at any point in the simulation via an *Enter* event. In health economic evaluations, it has been customary to create all entities at once at the beginning of the simulation but this need not be the case. Entities can enter later in the simulation due to incidence, births, immigration and other such mechanisms. Entities stay in the simulation (even if nothing is happening to them) until an *End* event occurs, at which time any final processing happens and they leave the simulation. Although death is the most common type of *End* event, entities may leave for other reasons such as end of their time horizon, loss of the indication, emigration and so on.

Entities are individualized by assigning to them specific values for their *attributes*. All entities share a common set of attributes (e.g., all would have an “age” if that attribute was part of a given model) but individual entities are assigned their own unique values for each attribute. These individual characteristics are updated throughout the course of the simulation, due to the natural flow of time, or to events that the entity experiences (e.g., medical complications, visits to medical professionals, change in compliance with prescribed treatment regimen), or due to other events affecting the entire system (e.g., a change in reimbursement, introduction of a new process) or even to the experiences of other entities (e.g., death of a caregiver). Aspects of the history of each individual can be retained as attributes if they affect the occurrence of subsequent events, or decisions about management, or other attributes, or any other aspect modelled. These attribute values are accessible to the model and can be output for display or further analysis. Needless to say, heterogeneity in the population, treatment
effect, and outcomes can be incorporated readily into the simulation as the attributes are individual.

In a discrete event simulation, time is explicitly modelled and it runs continuously as it does in reality. This means that the timing of events is unrestricted – they can even occur simultaneously (within the same day, for example) and competing risks can be properly applied. There is no need for “half-cycle” corrections or other devices to reduce the error introduced by fixed time cycles. It also enables the analyst to appropriately implement any time dependencies that exist and to keep track of the timing of events, the duration of conditions and so on. Delays — both deliberate (e.g., one day in hospital for a procedure) or implied by another condition (e.g., waiting for a bed to be available) — are readily incorporated. For economic models, the explicit handling of time permits full continuous discounting.

A particular strength of discrete event simulations is that resources are also modelled explicitly. They are considered special events where a resource is used. These resources can be people (e.g., doctors, nurses, therapists), goods (e.g., drugs, intravenous equipment), organs (e.g., liver transplants), space (e.g., hospital bed) or any other aspect that is conceptualized as a resource. The resources have unit costs which can be specified per “use”, per hour, and so on, and can these can be time-specific (e.g., more expensive during off-hours); even the costs of idle time can be specified. They also have units of capacity (e.g., Beds in an emergency room, doctors in a region). Thus, a type of resource can become a “bottleneck”, creating queues as individuals who need the resource wait for it to become available. If such resource limitations are a relevant feature of the decision problem in an IQWiG evaluation, it is recommended to choose a discrete event simulation rather than a Markov model.

Another feature of discrete event simulation is that individuals may interact with each other [3,6,74]. This can take many forms. For example, discrete event simulation allows for specification of constrained resources, with individual entities “competing” for those resources, and queuing up to use a resource when it is occupied. In health care, where constrained resources are ubiquitous, assessing how new interventions might free-up limited resources can be critical. Other examples include changes in behaviour. For example, as medical professionals gain experience with a health technology and see more patients, their
efficiency and subsequent treatment patterns for new patients may be influenced by their experience with previous patients. Finally, in the area of infectious diseases, where transmission of infections is a direct function of the infectious status of the individuals in the community, the ability to specify interactions and infection rules between entities is highly pertinent (see Section 10.2.5).

Discrete event simulations as well individual-level Markov models use equations of many types to control the modelling process. These can be linear functions, logistic regressions, hazard functions, Cox proportional hazards, or any other type of function that can be mathematically captured. A discrete event simulation also uses distributions like exponential, Weibull, beta, gamma, and so on to represent the variability in inputs, event times and so on. Together, these create the stochastic behaviour of the model. Random numbers must be generated by advanced algorithms that ensure sufficient “randomness” in the sequence.

Given the stochastic nature of discrete event simulation, any single run of the model may yield results that are far from typical. Thus, it is important to replicate the runs many times. Each replication uses a different set of random numbers and thus all the process that use random numbers, including the creation of the population and all of its experiences are unique. This yields a different set of results for each replication. Over many such replications, the result values will tend towards their “average”, providing a more stable estimate of the results. A common question is “How many replications are needed?” There is no single answer to this question as it depends on the variability in the model, the patients, and so on. The best way to judge it is empirical: examine the standard deviations of the critical results between replications. When the width is acceptable, there have been enough replications.

While it is possible to program discrete event simulations to act like cohort models — a temptation when data are limited — the technique’s flexibility allows for the simultaneous specification of multiple levels of complexity within the same model. In cases where data are highly scarce, the same model can be run with a simple set of inputs (e.g., by using average treatment effect instead of sampling from a distribution of treatment effects), but also in more sophisticated scenarios. This allows the analyst to address how the results might change if richer data were available to populate the model (e.g., what would happen if treatment effect is allowed to vary at the level of the individual).
Application of discrete event simulation and its use in health economics has been described elsewhere [3,54,56,58,64,81]. As discrete event simulation provides a framework for very natural representation of disease and resource use, it is well-suited to modelling cost outcomes related to health care interventions. Furthermore, modern software allows for a very transparent representation of the relationships in the model, for example, possible patient flows through the simulation. Software packages such as Arena® [74] allow for programming of discrete event simulations using flowchart methods very similar to those clinicians are already familiar with. They provide a visual representation of patient flows and possible outcomes. As discrete event simulations can produce a wide range of intermediate and final results, assessing the outcomes at numerous levels is possible. This is useful for validation purposes, and also provides a much more comprehensive set of results that can address the concerns of different decision makers and analysts.

### 10.2.3 Markov Models

Markov models are a powerful and rather simple tool for modelling prevention, diagnosis, and treatment of chronic diseases, where (1) parameters are time-dependent (e.g., restenosis after coronary intervention, relapse after cancer treatment, AIDS after initiation of HAART), (2) time-to event is important (e.g., the time to an event such as stroke, cancer relapse etc. matters), and (3) repeated events may occur (e.g., restenosis, late complications, second myocardial infarction) [82].

Several good accounts on how to construct Markov models are available [61,69,83], and the analysts is referred to these for more details. As noted above, Markov models represent reality in terms of a series of discrete states that reflect the conditions of interest (e.g., no disease, disease, death). In a cohort Markov model [83], the population is distributed into the existing states and then the model calculates the proportion of the cohort moving from one state to another (transitions) in a fixed period of time (cycle). (Though this limitation is not mandatory [84], Markov models in health care evaluations have been consistently implemented this way.)

States in a Markov model are mutually exclusive (no proportion of the population can belong to two states in the same cycle) and exhaustive (the sum of the proportions in each state must equal 100% in each cycle). From absorbing states, there is no exit: subsets of the cohort
entering that state remain there over the rest of the modelling period. Death is the most common absorbing state. Each state has a value according to the units of analysis (a “cost” in an economic model, for example).

In cohort models, the population in each state is defined by a common set of characteristics which describe that segment of the population (e.g., age, sex, disease severity) with no distinctions among individuals [58,61]. These can affect the likelihood of moving to other states in the next cycle (e.g., proportion with disease who die depends on age), and the value of the state (e.g., costs accumulated during each cycle depend on disease severity) [83]. The characteristics of each state can also change over time. This thinking in states and transition probabilities make Markov models appealing to some health scientists and physicians because this resembles disease stages (e.g., NYHA class in congestive heart failure, cancer stages). Moreover, data are often published in a way that facilitates direct use in Markov models. As mentioned above, however, the format of the available data should not lead the analyst to choose a Markov model over another model type.

In cohort models, updating of the characteristics defining each state may fail to take into account that these are altered by characteristic-dependent selective transitions out of the state. This can lead to substantial distortions (e.g., with one-year cycles, the mean age is typically increased by one year even though it is the older people who may be transitioning out due to death and disease). This aspect of cohort Markov models can be overcome by accounting for the heterogeneity in terms of additional health states that can be reasonably assumed to be homogeneous. For multiple patient characteristics, however, this can lead to a rapidly expanding set of states that soon becomes infeasible to implement. In this case, individual-level simulation of Markov models can be used to overcome the problem.

The simplest Markov models are chains that assume that the transition probabilities are constant over the entire modelling period [69]. Markov chains are a very attractive way to model because they can be solved numerically [85] by inverting the related matrices (i.e., it is not necessary to run the model at all). This property of constant transition probabilities is almost never realistic in health care, however. For example, background mortality rates of individuals change with increasing age. As spreadsheet packages and, more recently, dedicated software allow iterative solutions to Markov models which are not associated with the technical complexity of a matrix algebra solution, health economic evaluations tend to use
Markov processes that allow the transition probabilities to vary over time (e.g., for the risk of death to increase as the model progresses through time) [68].

Another aspect of Markov models relating to transition probabilities is the Markovian assumption [57,61,83], which states that the probabilities of moving from one state to another depend only on the originating state, and not any prior states, duration or other characteristics that the population in that state may have experienced in the past. This assumption is often referred to as the memoryless property: that is the model has no memory of what has happened to the patient before. Therefore, a sufficient set of Markov states that captures the important patient histories must be chosen to provide a realistic model of the disease.

Markov models handle time by specifying fixed cycles over which the transition probabilities are applied, with any segment of the population allowed only one transition per cycle. This requires additional techniques to properly handle competing risks [58]. The analyst must choose a sufficiently small cycle length to mimic the continuous time flow of reality.

In a pure Markov model, interactions between individuals are not considered [3,6,54]. Some software has meanwhile tried to relax this restriction by allowing “external” parameters such as disease prevalence to be used and updated in the model. This feature can be used to incorporate some transmission features in the Markov framework.

Individual-level Markov models, often referred to as Monte Carlo simulation because of the techniques used to sample probability distributions, overcome many of the limitations imposed by the cohort implementation [6,12,61,69]. In these models, individuals are assigned their own unique characteristics, their historical experiences can be retained, and heterogeneity in treatment effects and outcomes can be accounted for. Unlike cohort models, Monte Carlo simulations are stochastic, and all elements of the simulation can be specified using probability distributions. Patient flow through the model is determined by sampling from these probability distributions. A difficulty arises when these models cycle one individual at a time applying all the transition probabilities each cycle even if the individual experiences no transition in the majority of cycles. This slows processing time substantially.
Similar to discrete event simulations, individual-level Markov simulations can use equations of many types to control the modelling process (e.g., linear functions, logistic regressions, hazard functions, Cox proportional hazards, etc.)

Markov models can be modified further in various ways to overcome the limitations of a pure Markov model. For example, “tunnel states”, “half-cycle corrections”, multiple substates, external macros, and other features of available software are routinely used in practice. However, these features counter one of the main appeals of Markov models — their simplicity.

As with discrete event simulation, new software has simplified the programming of Markov cohort and individual-level simulations, allowing for intuitive visual representations of the model processes and facilitating transparency. In addition, there are pre-specified tools supporting the visualization of the state distribution over time (so-called Markov trace), sensitivity analysis, presentation of uncertainties, interface to spreadsheet software and macro language, and many other efficiency tools.

**10.2.4 Agent-based Simulation**

Agent-based simulation is a powerful further development of discrete event simulation, but even less commonly used as a modelling technique in health care [53,67,86-88]. The main difference compared with standard discrete event simulation is that in an agent-based model, the entities are aware of each other and of their environment (renamed *agents*) [67]. It is highly suitable for modelling infectious diseases where transmission of infection is important. Areas where social behaviour plays a key role in influencing the outcomes may also be relevant applications. While basic discrete event simulation can accomplish many of these things, agent-based simulation is a more natural approach, especially where geographic or spatial considerations are important. It facilitates the programming process as well as allowing for a more transparent presentation of how the model operates.

An agent, like an entity, will often represent patients, but may represent people who are at risk for the illness at issue. Like an entity, agents can also be specified to represent other things such as medical professionals, or even organizations and communities. Agents can modify their behaviours over the course of the simulation and influence the behaviour of other
agents, and thus one of the critical components of an agent is its “capability . . . to make independent decisions [and] . . . to be active rather than purely passive” [67]. Agent-based simulation also allows for spatial relationships between agents to be specified.

As with discrete event simulation, new software has simplified the programming of agent-based simulations, allowing for intuitive visual representations of the model processes and facilitating transparency [53]. These visual tools are not only useful for the decision maker but also for the analyst when debugging the program.

Some of the scenarios where use of agent-based simulation may be considered [53] are as follows:

- behaviour of individuals is non-linear and possibly best described by discontinuous functions (e.g., thresholds),
- individual temporal correlations, adaptation and learning need to be represented,
- interactions between individuals are heterogeneous,
- network effects may be present,
- fluctuations are possible in the system, with a steady state based on average relationships being unrealistic.

10.2.5 Transmission Models

In epidemiological and public health modelling of infectious diseases, transmission models are a much more common and accepted technique [60]. Transmission models [89] can be constructed at various levels of complexity and can be either deterministic or stochastic [55, 60]. In their simplest form, deterministic models are based on solutions to a series of differential equations which describe the infectious states (compartments) of the population (e.g., susceptible, infected, immune), the degree of infectiousness of the disease in question, and mixing patterns in the population. The effects of vaccination, for example, which alters the susceptibility status of the population and, potentially, the strength of infectiousness and other factors, can thereby be incorporated accounting for both the direct effects of vaccination (i.e., protection conferred upon the vaccinated individual), and its indirect effects (i.e., herd
immunity). Deterministic compartmental transmission models, which are in essence cohort models, however, require a number of simplifying assumptions which can be unacceptable in specific situations [55,60]. The simplest models assume complete uniformity in the population, instantaneous contacts and infection, instantaneous and “memoryless” mixing, uniformity in infection events, and uniformity in recovery from infection [60]. These assumptions often will not be supportable when providing information to decision makers about the economic consequences of related interventions. Some of the assumptions in transmission models can be relaxed with more sophisticated transmission modelling techniques. Techniques specifically designed to address those issues provide a more convenient and accurate modelling framework [60].

10.3 Recommended Technique

IQWiG does not recommend the use of any specific modelling technique. The critical consideration when determining the appropriateness of a modelling technique is whether it can accurately provide the information required by decision makers. This involves more than simply providing a point estimate for the expected economic consequences of different health care technologies — addressing the likely range of results as fully as possible is required. Transparency in modelling is also important, but it is the position of IQWiG that transparency should not trump in-depth modelling. With adequate documentation it should be possible to render any technique adequately transparent. It is incumbent upon the analysts to justify their choice of technique and software and assess the implications of this choice.

IQWiG also does not recommend which software or programming language should be used in modelling. However, the utilization of popular spreadsheet packages has to be avoided for modelling given the level of information needed, and the importance of the research questions being evaluated.

Finally it should be noted, that independent from the choice of the modelling techniques in some cases valid models can predict how these benefits identified in clinical trials appear under other circumstances or time perspectives. However, under no circumstances can models generate new benefits.
11 Handling Uncertainty and Variability

11.1 Types

There are two major kinds of variability in economic models [71]. There is variability that occurs naturally and that is of interest to the analyst and there is variability that one would prefer not to have but may be inevitable. The approach to these differs markedly. In the first instance, it is an object of study while in the second it is something to be eliminated as much as possible, with the residual quantified.

The costs may vary according to where the services are delivered, to what types of patients, and under what conditions, among other things. Thus, variability of interest has to do with characteristics of the programme, the patients, the context, and the unit costs. Note that these elements may be known with certainty (e.g., starting dose) or one may be uncertain about them (e.g., the per diem cost of a hospital stay). Either way, these determinants of the costs are relevant to the analysis and are, therefore, to be studied. This is done via sensitivity analysis, which can be univariate (change the value for one determinant of interest at a time), or multivariate. For the latter, it is most useful if it is done by carrying out analyses for specific scenarios defined by a set of values for these elements.

As part of the examination of variability of interest, the analyst should consider the decisions about the structure of the model and the assumptions made throughout. These may be related to the way the health-care intervention is conceptualized, the clinical events included, the time horizon used for modelling, the functional relations, the choice of costs and benefits to be addressed, as well as the methods of measurement and valuation of outcomes, among other things [34,90,91]. These elements may appear to be invariant to the modeller but there is usually disagreement amongst practitioners about the appropriate approach to be implemented some of these areas [92] (e.g., discount rate). If other plausible assumptions could be made for key elements of the model, then the resulting variability should be formally examined.

Variation may also result from statistical analysis of data sources (i.e., estimations from samples with finite sample size, which are necessarily never free from uncertainty), the modelling technique, or random individual variation. Thus, the costs may vary because there is uncertainty in some of their determinants such as the dose consumed over time (note that
the same element may have both types of variability); or because the model is stochastic (i.e., uses random numbers in Monte Carlo draws). This type of variability is not itself of interest and therefore various techniques are deployed to eliminate as much of it as possible. These variance reduction techniques (e.g., cloning) are dealt with extensively in the literature [59,93,94]. One common such technique is to increase the sample size by simulating more patients or running the model more times. Whatever variation remains after all the variance reduction techniques have been deployed must be quantified. This can be done through the use of various statistical methods applied to the model results or through analytic techniques such as probabilistic sensitivity analyses.

11.2 Quantifying Uncertainty

The models used in the economic evaluation of health-care interventions most often require synthesizing data from various sources, such as clinical trials, observational studies, data base analyses, surveys, and so on. Thus, there is some degree of uncertainty in all models [90] and the sources can be categorized in number of ways [91,95-97].

Parameter uncertainty arises from the fact that the value of some of the model parameters are estimated from samples with finite sample size. Thus, parameter uncertainty reflects our inability to determine these values precisely, such as those pertaining to the efficacy of the intervention, to its resource use consequences and to the valuation of those consequences [98]. This type of uncertainty is often described by means of confidence intervals or other statistical approaches reflecting the observed uncertainty in that parameter within the source data. As it is rarely practical to increase the samples in order to reduce this uncertainty around the parameter values, it is important to investigate its impact on the results.

The characteristics of the patient population, such as age, gender, relevant clinical history are very important in terms of the applicability of the economic evaluation’s results. Although the characteristics of the patients may vary across the population, this does not per se imply uncertainty [91]. There can be uncertainty, however, because of insufficient information on these characteristics, and if so, this should be quantified along with other parameter uncertainties.
Uncertainties in the model inputs have to be considered and their impact on the model results quantified using the proper methods [25,91,99]. The three main methods in use in health economic evaluations are univariate and multivariate deterministic sensitivity analysis and multivariate probabilistic sensitivity analysis. Other methods exist [93,100] but are rarely applied in health technology assessments.

Univariate sensitivity analysis addresses the range in outcomes resulting from varying one of the model parameters across its plausible uncertainty range, while all other parameters are kept at their original assigned values (it should be noted that this may not take into account correlations among inputs). This can give an idea of how large the impact of that uncertainty is, but clearly cannot quantify the full uncertainty in the results. Univariate sensitivity analyses are extremely helpful if a decision maker has assumptions regarding one important parameter that differ from the base case analysis.

Multivariate sensitivity analysis attempts to make the assessment of uncertainty more complete by changing more than one parameter at a time and investigating their joint impact. Although this can be very useful for examining scenarios when the uncertainty is of interest, it is time consuming when each parameter has a broad range and not very accurate if correlations among parameter values are not taken into account.

In probabilistic sensitivity analysis, each parameter that is subject to uninteresting uncertainty is assigned a distribution and these distributions are simultaneously sampled to create a set of inputs for one run of the model [34]. This is replicated many times and the resulting set of results provides an idea of the total impact of parameter uncertainty. For this to reasonably reflect uncertainty, it is important to choose distributions that accurately represent the uncertainty around the parameter estimates, and to correctly reflect how these correlate with each other. Otherwise, parameter combinations that rarely (or never) occur in actuality may be overrepresented.

11.3 Recommendation

Variability that is of interest to investigators should, of course, be investigated. Parameter and other uncertainty that is not of interest but irreducible has to be quantified. All analyses undertaken to do so should be documented fully, in terms of the ranges for the parameter
values used and assumptions made. It is recommended that univariate sensitivity analyses are not replaced but complemented by multivariate and probabilistic sensitivity analysis. For univariate and multivariate analyses, the results should be reported in both tabular form and as a “tornado” graph giving the range of results for the input range(s). In case of probabilistic sensitivity analyses, results should be presented as the cumulative distribution of results. Cost-effectiveness acceptability curves [101] are not requested in IQWiG evaluations, given that these analyses are not about threshold cost-effectiveness ratios. Structural sensitivity analyses should be undertaken to explore the impact of varying structural assumptions.
12 Validation

The approaches to validation have been classified in many ways. Rather than focus on a particular scheme and its terminology, it is important to understand what validation consists of and what each part covers. One key element has to do with whether the model makes sense. This face validity relates to the influence diagram, the model concept, the approach to the data, the functional relations and the choice of modelling technique. They must all be consistent with prior knowledge, best available information and best practices for each component. A second key element in validation is whether the model is implemented correctly (also called technical validity, or “verification”). This aspect has to do with whether the model is operating in the manner intended — that the logic of the “program” is correctly implemented, with a minimum of programming errors. The third key aspect is external validation of the model — establishing whether the model correctly reflects the “real world”, within the defined limitations of the model. This is extremely important since a technically perfect, bug-free model with strong face validity is still of little value for decision making if it does not accurately reflect the system and its behaviour.

A simulation model that is valid for one purpose may not be valid for another [102]. The external validation process must therefore cover all intended uses of the model, and if it is put to different purposes later, then renewed validation should take place.

There is no single procedure for validating a model, but there are some basic steps that have to be taken [103].

The design and structure of the model should be reviewed by other people with knowledge of the area the model is covering. This should be done for the influence diagram and model concept. The influence diagram is an important aspect of face validity. As it displays the modellers’ understanding of the illness and its management, it provides a clear tool for clinical and other experts to assess the planned model.

All equations to be used in the model should be tested separately before being entered into the model. This is done to ensure that they are correctly expressed and give the right results for entered values. The equations should also be calculated with extreme values on the borders of the model scope to make sure that there are no problems in those regions.
It is a huge task to test all aspects of the model when it is completed, due to the complexities and interactions. The only sensible approach is to keep testing the model as it evolves. This is best done by constructing the model in modular fashion. Each module is tested when it is developed. The module testing consists of checking that, given the inputs to the module (e.g., entities and data entering it in a discrete event simulation), its output is correct. This testing needs to extend to extreme inputs and conditions to ensure that the model runs correctly even in these cases. Any changes to a module already tested trigger new testing of that module. As the modules are added on, continuous testing should take place, thus ensuring consistent behaviour. In an individual-level simulation (Markov modelling with first order Monte Carlo Simulation or Discrete Event Simulation), a basic method is to test single, well-defined patients to make sure the paths and events he or she experiences are the expected ones. Moreover the same approach chosen to test the single modules should be used for the whole model.

A good way of externally validating a model is to run it using the inputs from one or more historical situations, such as completed clinical trials, registries or outcomes studies, from which data have not been used in the tested model. The model results are then compared with the known ones [9]. By testing trials that have not been used as sources for inputs, the ability of the model to reproduce reality is checked. This checking must be tempered, however, as trials themselves are substantial deviations from reality. Ideally, the same process should take place with data from actual practice, although these are rarely available.
13 Model Documentation

A detailed technical report describing all of the modelling steps from development of the initial influence diagram to final validation is required. In addition, a fully executable version of the model must be made available, along with a user manual on its operation. In line with other suggested guidelines [5,12,20,25,99], documentation on the model should include the following:

- The influence diagram used to guide model development

- Details of the model concept
  - Description of the population(s) considered in the evaluation, including subgroups
  - Description of the evaluated health technologies
  - Selection of model settings (simulation size, time horizon, discount rates, etc.) and justification
  - Review of economic evaluations in the therapeutic area in question

- Description of all data sources. Justification for selection of data sources must be provided.

- Details of all functional relationships used in the model. If they were custom developed for the model, the methods used must be provided in detail.

- Listing of all assumptions related to both data source and model structure. Especially important is a detailed account of any assumptions and techniques used to project beyond the source data.

- Rationale for the modelling technique adopted
  - Description of how the technique conforms with the mandatory features (Section 7)

- Outline of the validation techniques used and results
• Detailed results, including an assessment of the impact of:

  o Use in relevant population subgroups

  o Uncertainty in input data

  o If probabilistic sensitivity analyses are conducted, include specification of

    ▪ the probability distributions used and sources

    ▪ correlations among input parameters

    ▪ any structural variants

  o Interpretation of results, including a description of limitations of the approach adopted.

An electronic version of the model must be provided with the understanding that the model will enter the public domain, and may be adapted for use in future evaluations. The electronic model must be fully accessible and allow reviewers and the public to view all formulas and relationships included in the analysis and to run the model using different input data. To facilitate review of the model, the electronic model should be accompanied by a user’s guide describing the software and hardware required to use the model, how model inputs can be modified, where these inputs can be found in the model, and how the model may be run and results extracted.
References


102. Law AM, McComas MG. How to build valid and credible simulation models [Online]. 2001 [last access 23.06.2008]. URL: http://www.informs-sim.org/wsc01papers/004.PDF.