# THE ROLE OF MODELLING IN ECONOMIC EVALUATIONS IN HEALTH CARE

EDITED BY
BALÁZS NAGY
JONATHAN D. CAMPBELL
ZOLTÁN KALÓ

# THE ROLE OF MODELLING IN ECONOMIC EVALUATIONS IN HEALTH CARE

# THE ROLE OF MODELLING IN ECONOMIC EVALUATIONS IN HEALTH CARE

**BALÁZS NAGY**
**JONATHAN D. CAMPBELL**
**ZOLTÁN KALÓ (EDS)**

**BUDAPEST, 2019**

AFFILIATIONS:

**Editors**
Balázs Nagy – Eötvös Loránd University, Syreon Research Institute
Jonathan D. Campbell – University of Colorado
Zoltán Kaló – Eötvös Loránd University, Syreon Research Institute

**Authors**
Ahmad Fasseeh – Eötvös Loránd University, Syreon Middle East
Anett Molnár – Syreon Research Institute
Balázs Nagy — Eötvös Loránd University, Syreon Research Institute
Bertalan Németh – Semmelweis University, Syreon Research Institute
László Szilberhorn – Eötvös Loránd University, Syreon Research Institute

© Authors, 2019
© Editors, 2019

ISBN 978-963-489-119-2
ISBN 978-963-489-120-8 (PDF)

EÖTVÖS UNIVERSITY PRESS www.eotvoskiado.hu

Executive Publisher: The Dean of the Faculty of Social Sciences of Eötvös Loránd University
Cover: Ildikó Csele Kmotrik
Printed by: CC Printing Ltd

# CONTENTS

# 1 Introduction

Balázs Nagy

Imagine you are the head of your family and need to make the most out of your budget to keep your loved ones healthy. Now imagine yourself as the manager of a hospital who has to plan the medications and interventions the hospital should buy for next year. Let's say you are the chief of the national health insurance fund in your country and you need to make decisions about the interventions to reimburse and the healthcare provisions to promote. Finally, imagine you are the Minister of Finance in a country who decides whether the society will spend the tax revenues on building highways or on emergency care units. In all these scenarios, you have to take into account at least two things: how much money you can afford to spend (i.e. affordability) and what the return will be for any outlay of money invested (i.e. efficiency). Economic evaluations in healthcare are intended to answer these two questions.

The question of affordability is answered by the budget impact analysis which is used to estimate the likely change in expenditure to a specific budget holder resulting from a decision to reimburse a new healthcare intervention (or some other change in policy) at an aggregate population level (Mauskopf, Sullivan et al. 2007, Roberts, Russell et al. 2012, Consortium 2016). Beyond having an idea about affordability, there is another crucial economic question: how should the available money be spent the best way? The question of efficiency is answered by various types of cost-effectiveness analyses – or the so-called full economic evaluations[1] – which compare at least two healthcare programs (e.g. medications, procedures, investments) by looking at both costs and health outcomes that include benefits and risks (Gray, Clarke et al. 2010).

To meet the goals of both resource constraints and efficient allocation, healthcare analysts face a number of limitations:

- Data about the effectiveness of interventions is often sparse or limited.
- Clinical data is primarily designed to answer clinically, and not economically, meaningful questions. E.g. clinical trials target a very specific population, do not compare all relevant alternatives, do not encompass appropriate time horizons and do not provide information on economic outcomes.

1 Four types of full economic evaluations are distinguished in health care: cost-minimization, cost-effectiveness (according to the narrow definition), cost-utility, cost benefit analysis.

– In healthcare there is also uncertainty around the benefits of the treatment, length of impacts, real-life vs. artificial (clinical) settings, heterogeneity of information and context-specific variations in the results.

All these limitations need to be taken into consideration by the analyst. At the same time there is a strong pressure to buy as much healthcare and as early as possible. Patients can't wait long for decisions. There is a pressure on buyers to pay for the best available medications while their choices are limited by financial constraints. In the end some patients' needs have to be traded off against the needs of others.

From the supplier's perspective healthcare is a very resource and investment intensive industry. Research and development is costly, risky and time demanding. Pharma companies have the time only until patent expiry (appr. 20–25 years) to realize a full return on their R&D investments. Out of tens of thousands of molecules, only a few will reach the pharmaceutical market to accumulate large enough revenues to support successful business continuation.

In the end, when it comes to the allocation of limited resources, healthcare decision-makers face the pressure from both the public and industry side, where any mistake (e.g. delay in supporting cost-effective interventions or supporting non-cost effective investments) results in welfare loss to the society.

To facilitate early and efficient decisions while at the same time circumventing limitations due to time and information barriers, healthcare researchers often apply economic modelling. Since the early seventies these methods have gone through an immense improvement (Weinstein 2006). The rate of development seems never-ending with new methods and approaches emerging as the quality and quantity of data expands, as needs of decision-makers change and as statistical, mathematical and computation methods improve.

This piece of work gives you an overview of these techniques with regards to their usefulness in conducting full economic evaluations in healthcare. Specific terms and methods are systematically presented and discussed using the experience of the authors and other researchers in the field.

First the concept of modelling is presented. Then the architecture of decision models is discussed, after which the model building methods are described. In the 4$^{th}$ and 5$^{th}$ chapters handling uncertainty and validation methods of decision models are discussed.

# 2 What is a model?

Balázs Nagy and Bertalan Németh

The term 'modelling' is broad in itself and specific definitions are linked to specific contents. Still János Neumann's general definition gives the best idea of what modelling is: "The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work – that is correctly to describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain esthetic criteria – that is, in relation to how much it describes, it must be rather simple." (Bródy and Vámos 1995).

In brief, models are intended to be the simplified representation of real-world situations that help answer specific research questions. They are to remain as simple as possible while retaining the details necessary to approach the specific question (Group 2010). There are a number of techniques called models in healthcare, and modelling is not restricted to one specific method or approach. In the end models cover the process which combines techniques and skills of mathematics and computation to steer people in need to the right direction in order to answer questions or make decisions.

A number of things can be modelled in healthcare. For example, models can help us forecast events or help us relate one concept to another including (Group 2010):

i) future supply and demand
ii) links between demographic and other factors
iii) patient health behavior
iv) healthcare access
v) spread of communicable diseases
vi) optimal healthcare delivery.

Health technology assessment (HTA[2]), and within HTA, economic_evaluation has placed some very clear requirements on researchers in terms of conducting proper analyses.

2 HTA is a multi-disciplinary field that addresses the clinical, economic, organizational, social, legal, and ethical impacts of a health technology, considering its specific health care context as well as available alternatives International, H. T. A. (2017). What is HTA? Accessed Aug. 1, 2017. HTAi.

These include the need to incorporate all appropriate evidence into the analysis, to compare new technologies with the full range of relevant alternative options and to reflect any uncertainty in evidence in the conclusions of the analysis (Briggs, Claxton et al. 2006). The need to satisfy these requirements and overcome the issues we discussed in Chapter 1 provides a strong rationale for using decision analytic modelling as a framework in economic evaluations.

## 2.1 Decision analytic modelling

Decision analysis is used to construct and structure decisions in many areas of the economy. It includes multiple methods and tools to identify, represent and assess the important aspects surrounding a decision – not only in healthcare but in a wide range of disciplines such as marketing law and engineering. Decision analysis for the purpose of economic evaluation in healthcare is a "systematic quantitative approach to decision making under uncertainty where at least two decision options and their respective consequences are compared and evaluated in terms of their expected costs and expected outcomes" (Gray, Clarke et al. 2010).

Decision analytic models "use mathematical relationships to define a series of possible consequences that would flow from a set of alternative options being evaluated. Based on the inputs into the model the likelihood of each consequence is expressed in terms of probabilities and each consequence has an expected cost and an expected outcome." (Briggs, Claxton et al. 2006). These models can serve a number of purposes:

- structure the research questions,
- synthesize evidence,
- extrapolate beyond observed data,
- link intermediate and final endpoints,
- generalize results to other settings of patient groups,
- demonstrate uncertainty around the decision and
- indicate the need for and value of further research.

They usually do not provide straight ‚yes' or ‚no' answers, but a framework for decisions. The key purpose of decision modelling is to allow for the variability and uncertainty associated with each decision. E.g. what are the expected costs and benefits of introducing a nationwide diabetes screening program for people between age 25–65, or is it worth building an outpatient care unit in a distant town with only 10,000 inhabitants?

## 2.2 Taxonomy of decision models

A comprehensive taxonomy of decision models in healthcare[3] was provided by Brennan and colleagues in 2006 (Brennan, Chick et al. 2006). Their paper gives an overview of model types being used for economic evaluation in healthcare. This section is based on their study. Table 1 shows the range of existing modelling approaches for healthcare technology assessment. This taxonomy divides methods according to 6 dimensions: time, interaction, cohort-individual level, expected values/being Markovian, randomness, and heterogeneity. Each grid cell in the table is related to its neighbors by varying some of the basic assumptions that underlie each model type. **Rows 1–4** describe factors involving both time and interaction between individuals. The most commonly used healthcare modelling approaches are largely those in the top half of the table: models without simulating interaction between individuals, and with (**row 2**) or without (**row 1**) explicitly modelling the passing of time. Models with interactions (**rows 3, 4**) are important when individual interactions are influential (e.g. to understand the progress of diseases as in the case of infectious diseases transmission in the population) or there are constraints which affect individuals (e.g. finite service capacity or restricted supplies of organs for transplantation). For these categories (**rows 3, 4**) discrete time and continuous time models are distinguished.

**TABLE 1 TAXONOMY OF MODEL STRUCTURES FOR ECONOMIC EVALUATION OF HEALTHCARE TECHNOLOGIES**

| | | | | A | B | C | D |
|---|---|---|---|---|---|---|---|
| | | | | Cohort/aggregate level/counts | | Individual level | |
| | | | | expected value, continuous state, deterministic | Markovian, discrete state, stochastic | Markovian discrete state individuals | non-Markovian discrete state individuals |
| 1 | | no interaction allowed | untimed | decision tree rollback | simulated decision tree | individual sampling model: simulated patient level decision tree | |
| 2 | | | timed | Markov model evaluated deterministically | simulated Markov model | individual sampling model: simulated patient level Markov model (variations as in quadrant below for patient level models with interaction) | |
| 3 | | interaction allowed | discrete time | system dynamics (difference equations) | discrete time Markov chain model | discrete time individual event history model | discrete event simulation |
| 4 | | | continuous time | system dynamics (ordinary differential equations) | continuous time Markov chain model | continuous time individual event history model | discrete event simulation |

Source: based on Brennan et al. (2006)

3 Here models with the purpose of examining the cost-effectiveness of healthcare interventions are in mind.

**Columns A–D** entail cohort and individual level models and disentangle assumptions concerning expected values, randomness, and the heterogeneity of entities. Cohort models in **columns A, B** take into account the proportion of people with common characteristics. They can distinguish different attributes (e.g. ages, weights, genders, stages of natural history of disease, or other risk factors) by subdividing the number of states or branches. As the number of dimensions rises exponentially (e.g. binary attributes imply the duplication of dimensions), these cohort models are restricted in their ability to model complex situations (e.g. diseases with multiple complications, patients with a long and complex disease history, or the progression and treatment of multi-stage diseases, (e.g. rheumatoid arthritis, types of cancer). In many cohort models the Markovian property is typically assumed, meaning that the future is conditionally independent of the past (see more in section 3.2). Appendix I provides detailed explanation about each grid cell.

In Appendix II an example of using 4 different modelling techniques for the same disease and intervention area (pertussis immunization) is presented. Model "A" presents a decision tree model, Model "B" presents a Markov model, Model "C" presents a discrete event simulation (DES) model and Model "D" presents a dynamic state transition model. This example illustrates how different models are applied to resolve decision problems in the same disease area. All 4 models focus on different levels and depth of problems using various assumptions on the comparator of interest, disease incidence, time horizon, herd immunity and other variables. In the chapter the most commonly used model types will be discussed in more depth.

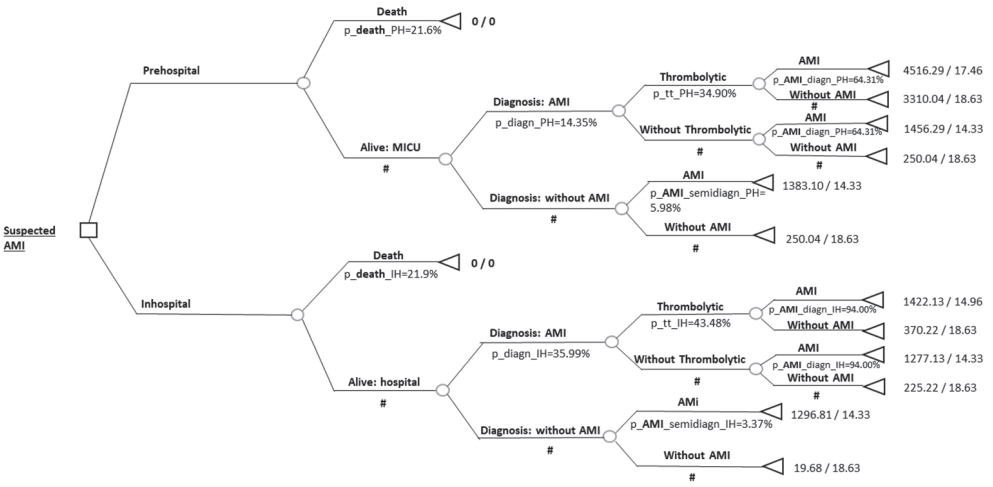
# 3 Architecture of decision models

Balázs Nagy, Anett Molnár and Bertalan Németh

## 3.1 Decision tree cohort models

Decision tree models use a tree-like structure to present decisions and their possible consequences. They identify alternatives and specify the sequence and linkage of events by using a branching structure in which each branch presents an event that may take place in the future. The consequences related to each decision take account of the probability of event outcomes, resource costs, and health consequences (e.g. life years, QALYs) (Gray, Clarke et al. 2010 pg. 188.). See an example of a decision tree in Figure 1.

**FIGURE 1 EXAMPLE OF A DECISION TREE MODEL STRUCTURE TO ANALYZE THE COST-EFFECTIVENESS OF PRE-HOSPITAL COMPARED TO IN-HOSPITAL THROMBOLYSIS IN PATIENTS WITH ST-ELEVATION MYOCARDIAL INFARCTION (STEMI) IN THE PUBLIC HEALTH SYSTEM**




Legend: AMI – acute myocardial infarction; p – Probability; PH – Prehospital; IH – In-hospital; MICU – Mobile Intensive Care Unit; # - 1 – the other probability. Outcomes are "cost (R$)/ life year"

Source: adapted from Araújo et al. (2008)

Decision trees are usually constructed from left to right, starting with the decision node (see the small square on the left in Figure 1) and ending with the outcomes. They follow the logical structure of the decision problem by tracking the sequence of events (see Figure 1). Any events that follow the decision are chance events and are characterized by probabilities. These events are presented by chance nodes in the decision tree diagram (see the circular symbols in Figure 1). Each outcome from each chance node is denoted by a line (branch) attached to the chance node and labelled (e.g. 'Diagnosis: AMI' on Figure 1). The likelihood of the event is represented by the branch probability (e.g. in Figure 1: 'p_death_PH' = probability of prehospital death = 21.6%).

The events stemming from a chance node must be mutually exclusive, hence the event probabilities should sum up to exactly 1. The order of events makes no difference in terms of calculating the expected value of examined strategies, however, it can have implications for how easy it is to perform the sensitivity analysis (see more in section 5.2) and to deal with complex treatment and/or disease pathways. The final outcomes from the alternative decision tree pathways end in terminal nodes (represented by triangles, see Figure 1). Each terminal node has cost and health outcome values or payoffs assigned to it (e.g. on the arm 'Inhospital' → 'Alive: hospital' → 'Diagnosis: AMI' → 'Thrombolytic' → 'AMI' the cost is 1422.13 and the QALY is 14.96). Payoffs include the costs related to the events in the decision tree and the final outcomes (life years, utilities, QALYs) and are presented in the model diagram at the terminal node (right of the triangles in Figure 1).

Once all the probabilities and payoffs are entered in the model it is possible to perform the analysis. Modelers often say that decision trees are 'averaged out' and 'folded back' (or 'rolled back'), which means that by folding back the tree the expected values of each strategy can be calculated. The folding back process starts at the right side of the tree and then averages back. As shown in Figure 1 the payoffs at each terminal node can be presented in costs and life-years gained which should then be multiplied by the probability of events taking place to arrive at any specific terminal node (i.e. as a result of the folding back process). The expected value of a decision is computed analytically by multiplying the probability of each outcome with its payoff and then summing the terminal node results related to the decision. In other words, the weighted average of payoffs for each strategy is summed and compared with other strategies' average payoffs to finally decide on the preferred strategy.

For example, in Figure 1 the cost of choosing in-hospital thrombolysis is calculated as:

(1-0.219)*0.3599*0.4348*0.94 *1422.13R$ = [0.1149] *1422.13R$
(1-0.219)*0.3599*0.4348*(1-0.94) *370.22R$ = [0.0073] *370.22R$
(1-0.219)*0.3599*(1-0.4348)*0.94 *1277.13R$ = [0.1493] *1277.13R$
(1-0.219)*0.3599*(1-0.4348)*(1-0.94) *225.22R$ = [0.0095] *225.22R$
(1-0.219)*(1-0.3599)*0.0337 *1296.81R$ = [0.0168] *1296.81R$
(1-0.219)*(1-0.3599)*(1-0.0337) *19.68 R$ = [0.4831] *19.68 R$

390.32 R$

Decision trees have to meet certain criteria:

- disease and treatment should be described with mutually exclusive patient routes,
- transition of patients to different routes should be based upon well-defined event probabilities,
- timing of patients' transition in their routes is not considered important or no timing of major clinical events within a route have relevance,
- each determined pathway should result in well-defined costs and clinical outcomes.

In decision trees, recursion or looping is not possible. This means that when decision trees are used to model diseases with lengthy prognoses or events that are likely to recur over time (e.g. in the case of chronic diseases) the structure does not permit movement back and forth between disease states. For example, the model depicted in Figure 1 is only able to simulate the short-term consequences of ST-elevation myocardial infarction, and recurrent events (e.g. next infarction or staying healthy for a while) are difficult to handle. In principle, such problems could be addressed by adding additional branches and extending the time horizon of the model. But as a consequence, complicated scenarios with many alternatives will manifest in long sequences of chance nodes and multiple outcomes, in which case the model can quickly become an unmanageable 'bushy' tree (i.e. many branches).

There is no implicit time variable within decision trees: the passage of time is accounted for by the outcome measures or payoffs. Implementing time dependency into a decision tree model can be difficult. For example, in Figure 1 whenever another infarction occurs it will be considered at the same time as the first one. This has not only impact on estimating the final outcomes in terms of adjusting the quality-of-life for the appropriate survival time, but also on ensuring the appropriate discounting of the value of both the costs and outcomes.

It is no surprise that decision trees are mostly suited to situations where events occur over a discrete short time period. These models provide a simple way to help identify strategies and their most likely manifestations/consequences/outcomes. Due to their design, they are also of great value in clarifying complex decisions.

As a general rule decision trees are mostly used for cases when:

i) there is short time horizon of disease, or time is not important for the analysis,
ii) only few and simultaneous events occur,
iii) simple back-on-the-envelope analysis of novel interventions are initiated and one needs quick results,
iv) one needs to stratify multiple choice decisions and wants to "see" the problem or the alternatives of a decision,
v) one needs to weigh up risks/benefits in a simplified way,
vi) one needs to analyze extreme cases/effects.

Other modelling techniques such as Markov modelling (as we shall see in section 3.2) can handle complexity and longevity in a better way. Finally it should be noted that decision

trees can often be used as subsets of larger models. For example, a decision tree can be built to identify the number of cases detected by a screening program which is then followed by a Markov model to estimate future costs and effects following detection (see more on hybrid models in section 3.7).

## 3.2 Markov cohort model

Markov models were named after a Russian mathematician (Andrey Andreyevich Markov) who introduced the term "Markov chain" in 1906 (Basharin, Langville et al. 2004). Markov chain is a random process that undergoes transitions from one state to another on a state space. The process is characterized by the so-called 'memoryless property', whereby the probability of a given transition in the system is independent of the nature or timing of earlier transitions (Drummond and McGuire 2001). In other words, Markov models work on the assumption that the future state of the object is determined by a random process dependent only on the current state of the object. This assumption is so basic to the methodology of Markov models that it is generally referred to as the 'Markovian assumption' (Group 2010). Markov models currently dominate the healthcare literature but these methods are also widely used to model non-healthcare related real-world processes.

Markov models, specified to healthcare, place patients (or other entities) into discrete 'health states', and time is partitioned into discrete periods, known as 'cycles', during which patients are assumed to stay in the same health state. An individual can be in only one health state during one cycle. In each cycle, a patient's health state may change from the current state to another health state (i.e. state-transition modelling) concluding each cycle in a finite number of states according to probabilities (Sonnenberg and Beck 1993). For each cycle, rewards are assigned to each health state and are earned at the end of the cycle. Rewards (analogous to payoffs in decision trees, see section 3.1) are expressed in costs, life years, quality adjusted life years or other types of healthcare/policy relevant outcomes.

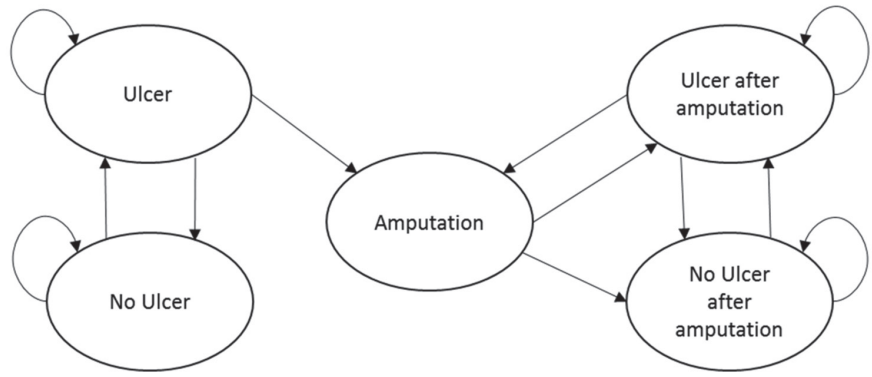
Markov models can either be described with

i) transition probability matrices or
ii) state transition diagrams or
iii) repetitive decision tree structures, as shown in Appendix III.

The transition probability matrix, as well as the state transition diagram concisely describe the potential state changes graphically. The repetitive decision tree structure may look needlessly complicated, though it is very helpful when a transition is being calculated through a series of event probabilities. Transition probability matrices use transition probabilities per cycle for patients in the cohort to change to another state. The rows of the transition matrix must add up to one (i.e. probabilities of moving from one health state need to add up to one).

**FIGURE 2 TWO WAYS TO SPECIFY A MARKOV MODEL: A) TRANSITION PROBABILITY MATRIX. B) STATE TRANSITION DIAGRAM**

A




B

| | No ulcer | Ulcer | Amputation | No ulcer after amputation | Ulcer after amputation | SUM |
|---|---|---|---|---|---|---|
| **No ulcer** | 0.9996 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 1.000 |
| **Ulcer** | 0.0223 | 0.9679 | 0.0098 | 0.0000 | 0.0000 | 1.000 |
| **Amputation** | 0.0000 | 0.0000 | 0.0000 | 0.9996 | 0.0004 | 1.000 |
| **Ulcer after amputation** | 0.0000 | 0.0000 | 0.0098 | 0.0223 | 0.9679 | 1.000 |
| **No ulcer after amputation** | 0.0000 | 0.0000 | 0.0000 | 0.9996 | 0.0004 | 1.000 |

Source: based on Tesar et al. (2017)

To process a Markov model, the model is run through a series of cycles and patients are redistributed in each cycle. In 'incidence models' all patients start from the same health state whereas in 'prevalence models' all patients begin each cycle distributed across health states. Then in the so-called 'cohort simulation'[4] the transition of the cohort among health states is followed from one cycle to the next depending on the transition probabilities. This technically involves multiplying the proportion of the cohort ending in one state by the relevant transition probabilities attached to that state in order to calculate the proportion starting in the next state. The simulation process results in the 'Markov trace' which shows the actual pathway of patients in the model (see later in section 3.4, Figure 5). The rewards (costs and outcomes) for each health state are accrued for each cycle and accumulated

4 The 'individual level simulation' differs from 'cohort simulation' in the sense that individual patients (instead of a cohort of patients) are directed through the model and, based on the transition probabilities, they are randomly moved from one state to another – this process is repeated a number of times (e.g. five thousand times) to give a robust result on the expected pathways of patients (see more in section 3.4).

through the entire length of the model. In the end, the accrual of costs and benefits are determined by the number of cycles and the proportion of the cohort that reside in each state over the time-horizon of the economic model. The cohort simulation provides the expected costs and expected outcomes for all examined strategies (e.g. the intervention arm and no-intervention arm) and it is possible to calculate the incremental cost-effectiveness ratio (ICER)[5] of implementing an intervention.

Markov models have clear advantages over decision trees in situations where timing of events is important and when events may happen more than once and a sequential or repetitive nature of events is important. Instead of possible consequences over time being modelled with a large number of possible pathways, as in decision trees, the disease progression is reflected as a set of possible transitions between the disease states over a series of discrete time periods (Gray, Clarke et al. 2010). In particular, Markov models are suited to modelling long-term outcomes where costs and effects are spread over a long period of time. Therefore Markov models are particularly suited to chronic diseases or situations where events are likely to recur over time (Gray, Clarke et al. 2010).

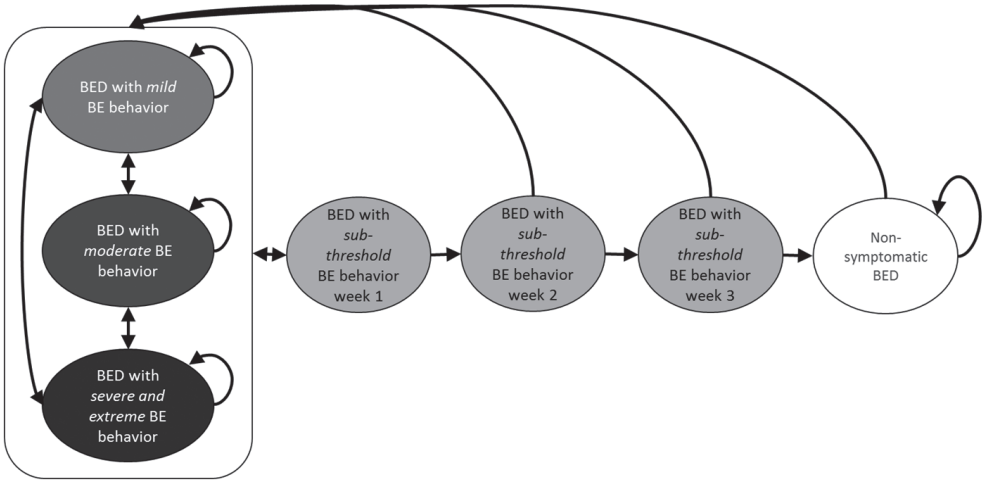
As a result of the Markovian assumption, these models are "forgetful", i.e. knowledge of the past is not required to predict the future. Many people believe that the Markov assumption causes Markov models to be extremely limited in application (Group 2010). For example, a person's probability of gaining weight is partly dependent on their current weight, but also partly dependent on their history of weight gain.

Nonetheless, it is, to some extent, possible to build memory into a Markov model. One can create new states that incorporate the memory for the desired trait. For example, in Figure 3 patients in a Markov model without any binge eating episode during the last week are assigned to health state 1 and if there is still no episode after 2 weeks then patients are assigned to health state 2 and after the 3rd consecutive symptomless period they reach the non-symptomatic health state.

It is also possible to incorporate time dependency into transition probabilities (e.g. patients have a higher chance of death as their age progresses). Models with changing transition probabilities are called 'process models', while models with fixed transition probabilities are called 'chain models'. Most healthcare models are 'process models' since death is a function of age and age changes as the model time progresses, thus, changing the transition probabilities to death.

5 ICER (Incremental Cost Effectiveness Ratio) is used to summarize the cost-effectiveness of a health care intervention. It is defined by the difference in cost between two possible interventions, divided by the difference in their effect.

**FIGURE 3 STRUCTURE AND PATIENT PATHWAYS OF THE COST-EFFECTIVENESS MODEL FOR THE TREATMENT OF BINGE EATING DISORDER**




Source: Ágh et al. (2016)

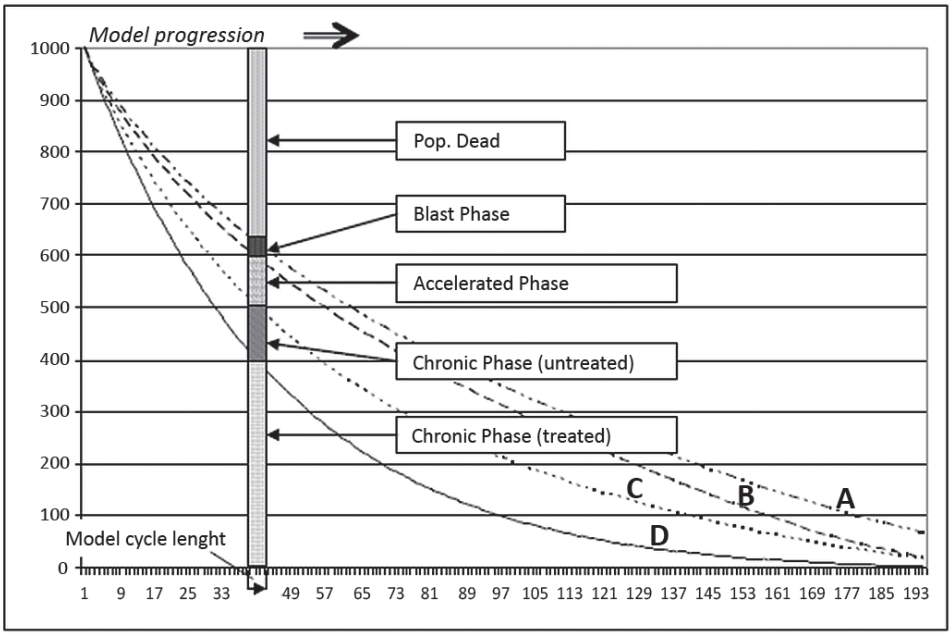
Markov cohort models are relatively simple to implement but can still simulate surprisingly complex systems. As such they are often a good first choice for modelling problems. However, for more complicated disease and treatment structures, the models may become very complex, especially in cases where the list of possible health states increases, and the manner in which patients move from one state to another are more difficult to track. Modelling complex chronic diseases where disease history matters, like diabetes, rheumatoid arthritis and schizophrenia, may result in an extremely large number of possible health states, outcomes and scenarios. For these cases, more sophisticated and flexible model structures may be preferred (see sections 2.2, 3.4 and 3.5 and Appendix I).

## 3.3 Decision analytic survival model

There is a special subset of state transition cohort models (see models in column A–B and row 1–2 in Table 1 in section 2.2), which is rarely mentioned as a distinct category: decision analytic survival models.

This type of modelling approach is especially common to analyze cases in which disease progression can be described by a stepwise sequence of health status deterioration (e.g. cancer treatment strategies) (Tappenden, Chilcott et al. 2006). In these models the average health status of patients is continuously deteriorating: patients move from one state to another without the mid- and long-term ability to improve.

**FIGURE 4 AN ILLUSTRATIVE EXAMPLE OF A DECISION ANALYTIC SURVIVAL MODEL WHERE THE COHORT PROGRESSION IS TRACKED WITH THE "AREA UNDER THE CURVE APPROACH"**




Key: A area = overall survival; B area = time in Chronic Phase plus time in Accelerated Phase; C = time in Chronic Phase; D=time in Chronic Phase and on treatment. Y axis = number of patients; X axis = time in the model (e.g. weeks)

In the example from Figure 4 chronic myeloid leukemia is modelled. The disease starts in the treated chronic phase, then patients move to the chronic untreated phase, then to the accelerated phase, then to blast phase and finish at death. Literally, there is hardly any chance for long-term improvement, i.e. the patients' average health statuses are continuously deteriorating (although there might be differences on the individual level). Since 'one-way traffic' is being modelled the analyst is permitted to simplify the state transition modelling framework: the pathway of the cohort is characterized by series of survival curves which determine the proportion of patients for each health status at any point. To estimate the proportion of patients in each health status and to associate cost and benefits to health states, the analyst has to simply compare the 'area under the curve' for each survival trajectory at the points of interest (e.g. 1st week, 2nd week, 3rd week, see Figure 4).

As long as appropriate empirical survival data is available, decision analytic survival models provide a convenient way to model continuously progressing chronic diseases. At the same time such structures hardly allow any flexibility when it comes to modification, amendment or extension of the structure (e.g. with alternative health states or treatment scenarios).

## 3.4 Markov simulation model

Decision trees and Markov cohort models, while being extremely useful to simulate a number of situations in healthcare, lack features which can be essential to mimic complex situations; for instance, the progression of a long-term chronic disease with multiple comorbidities, or consequence of changes in the provision patterns in a nationwide healthcare system. Markov cohort models can be impractical and may have difficulties handling:

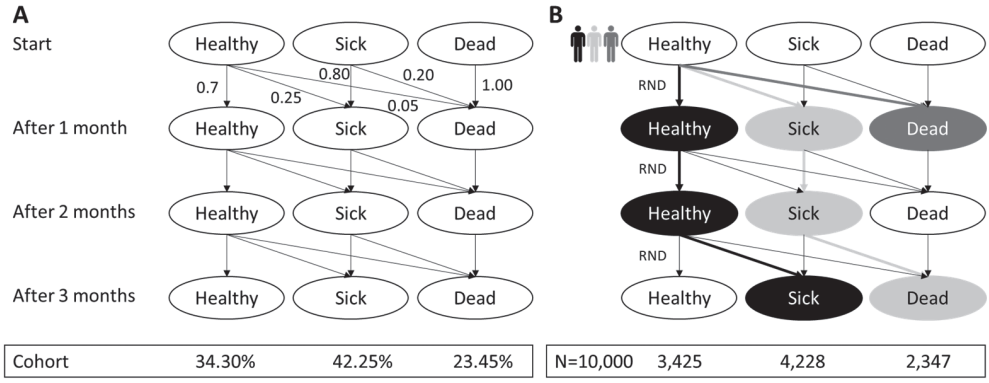
- memory: patients'/objects' behavior depends on their history which is difficult to track
- complicated cases: multiple complications call for many combinations of health states
- simultaneous or interrelated events: when multiple events occur together or when one event instantly leads to other mutually exclusive health states
- differences within heterogeneous patient groups: estimation of patient pathways and outcomes for subgroups of patients with different characteristics

To resolve these problems, patient level simulation models (also referred to as individual patient sampling or microsimulation models) are applied in practice. These models, instead of progressing cohorts of patients, simulate them separately and keep track of each individual's history. The simulated individuals can have heterogeneous characteristics which can alter their pathways in the model.

The simulation process starts by generating or selecting a group of individual patients with baseline characteristics (such as HbA1c, SBP, age, sex). The individual patient then passes through the model and when a decision node is reached, the pathway taken is determined according to the associated probabilities and a generated random number. All probability values and random numbers range between 0 and 1. When the random number is smaller than the probability value, the model assigns disease progression and vice versa.[6] The path followed by different patients will differ due to chance (see Figure 5). This process is called the 'Monte Carlo simulation'; it is also referred to as 'random-walk'. The model results in large numbers of simulated patient histories which are aggregated to provide the final results. The samples are expected to be large enough to successfully shrink the variability (due to "random walk") around the model estimates.

6 This is the logic for carrying out "random walk", however, there are other techniques to account for randomness in simulation models.

FIGURE 5 COMPARISON OF COHORT (A) AND INDIVIDUAL SIMULATION (B) IN A MARKOV MODEL




As patients in these models are tracked individually, it is possible to reflect on patients' characteristics and the history of their event. Also patients' characteristics can be updated over time as appropriate, their excess risk can be recalculated when necessary, and any number of competing risks can be simultaneously applied. When events happen, patients' characteristics can be updated. Time dependencies can be considered, and what happened earlier with patients can be tracked and stored for further use. There is no need to work of an average patient or restrict the analyses to homogeneous populations or run series of sensitivity analyses on different subgroups (Caro, Möller et al. 2010). Multiple comorbidities depending on multiple attributes can be modelled, while the number of health states can be greatly reduced and still real life circumstances can be accurately presented.

There are disadvantages as well to using individual simulation models. First, they usually demand significantly more data than cohort models. If various aspects of patient history are used to determine future prognosis, the model will require input parameters contingent on these patient characteristics. Second, the computational burden of these models is usually more than of cohort models. Robust model outcomes require a large number of patients to be simulated individually, which may be time-consuming. Depending on the complexity, the programming language and the pc infrastructure, there is a large spectrum in the running time, ranging from a couple of seconds to weeks. For example, the extremely complex Visual Basic for Applications (VBA) programmed Syreon Diabetes Control Model (Nagy, Zsólyom et al. 2016) would run for a week with 20,000 patients, while an 8 state VBA programmed model on schizophrenia (Németh, Molnár et al. 2017) with 20,000 patients would run for approximately 2 minutes. Third and importantly, individual simulation models have limited flexibility to analyze uncertainty.[7] Both deterministic and probabilistic sensitivity analyses (see more about these methods in section

7 Section 5.2 will outline the key concepts of uncertainty analysis in decision modelling.

5.2) are quite time consuming for individual simulation models. For a model with 10,000 patients a deterministic sensitivity analysis multiplies the number of runs by the number of variables of interest. The same model in the case of a probabilistic analysis requires two levels of simulation: one level based on fixed parameters to estimate a single expected value (first order uncertainty, see section 5.1); and a second to do sampling from a distribution of possible input values to assess uncertainty (second order uncertainty, see section 5.1). For the two levels of simulations, this would result in 100 million (10,000 x 10,000) individual simulations. This is only likely to be feasible for smaller patient level simulation models implemented in fast performing PCs and written in a 'simulation-efficient' programming language.

It is clear that despite their advantages patient level models are not always superior to cohort models. When a modelling exercise can be sufficiently carried out with the cohort approach patient level simulation is not encouraged. Markov cohort models are widely accepted by decision-makers, and only when these models reach their limits are Markov individual simulation techniques advised as a next step forward, which is usually the case when:

- complex disease and treatment pathways are to be analyzed;
- patients can develop different complications simultaneously;
- individual risk varies among patients;
- enough data are (or will be) available to populate the model;
- not all data are available, but the structure of the problem necessitates a complex modelling approach;
- the existing structure is potentially extended/complicated in the long-run;
- the analyst has good programming skills to execute the model;
- the model, in spite of its great complexity, can still be kept transparent and valid with all assumptions remaining transferable.

It is important to note that Markov cohort and Markov individual simulation models do not differ much in their structure (see Figure 5) and they actually have the same logic. As a matter of fact, Markov simulation models can be regarded as the extension of the cohort models with added variability and flexibility through the use of individual patient characteristics and the incorporation of patient history. Table 2 helps us understand the differences between Markov cohort and Markov individual simulation models and provides a good example on the choices the analyst has to make when considering state transition modelling.

**TABLE 2 COMPARISON OF THE FEATURES OF COHORT AND INDIVIDUAL LEVEL MARKOV SIMULATION MODELS**

| | Markov Cohort | Markov Individual Simulation |
|---|---|---|
| **Building time** | disproportional increase with model complexity | proportional increase with model complexity |
| **Data collection** | both types can be built using the same input data | |
| **Experience in use** | widely used | infrequently used |
| **Simulation time** | only needed for PSA | needed both to process the model and run sensitivity analyses |
| **Patient heterogeneity** | can rather be handled with additional formulas (increasing complexity) or with analysis of subgroups | can be handled with defining cohort or using patient level data |
| **Memory** | handled through adding tunnel states | handled through adding tracking variables to individuals |
| **Real-World/ Construction validity** | limitedly applicable | highly applicable |
| **Interaction due to co-variates** | limitedly applicable | highly applicable |
| **Timing of events** | adjusted to cycle length | |
| **Transparency/ Validity** | Transparent but if complex more difficult to validate | Transparent if interim results are provided |
| **Flexibility** | Limited in expanding the model with new data/assumptions | Unlimited in expanding the model with new data/assumptions |

## 3.5 Discrete event simulation model

In state-transition models (i.e. Markov models as shown in section 3.2 and section 3.4) the world is conceptualized as a series of snapshots using mutually exclusive health states. These snapshots are reflections of a fixed time period (i.e. cycle). In case of greater disease complexity, the analyst often has to increase the number of health states or reduce the length of the cycle and this (even in patient level Markov simulation models) could end up in too large, unmanageable (and/or even imprecise) models. Moreover, in Markov models with little probabilities to move from one state to another, needlessly large amounts of computations must be executed unnecessarily (i.e. when a patient does not have an event over a 5-year period, the model runs excessively from cycle to cycle for five years, which takes up unneeded computation time). Therefore it might be more useful to step out of the constraints of state-transition modelling and conceptualize the world in terms of consecutive events.

In discrete event simulation (DES) models the experience of individuals is modelled over time using the events that occur and the consequences of those events (Caro, Möller et al. 2010). Individuals undergo various events that affect their characteristics and outcomes. The term "discrete" refers to the fact that DES moves forward in time at discrete intervals (i.e., the model jumps from the time of one event to the time of the next) and that the events are discrete (i.e., mutually exclusive). These factors give DES the flexibility and efficiency to be used over a very wide range of problems in healthcare (Karnon, Stahl et al. 2012).

The most important terms to characterize DES models are as follows: *entities*, *attributes*, *events*, *resources*, *queues*, and *time* (Karnon, Stahl et al. 2012). Entities are objects that have attributes and consume resources while experiencing events, but consumption is not affected by individual-level behavior. Attributes are features or characteristics unique to an entity. They may change over time or not. An event is something that happens at a certain time point in the environment affecting resources and/or entities. Resources are objects that provide a service to an entity.[8]

DES models do not only permit a flexible individual-level analysis but are also useful tools to analyze processes at the population level. For doing so, using 'queues' is a key concept. In models queues are applied when several entities compete for constrained resources (Berger, Bingefors et al. 2003). A line structure enables interaction between entities to take place with constraints, and as such it enables a schedule of within- and between-patient events to occur throughout the modelling process. This allows the efficient processing of events as they happen throughout the population. This technique is not only close to real life circumstances but substantially reduces the calculation time: models can usually consider individuals simultaneously while the 'model time' is permitted to jump to the occurrence of the next event rather than proceed in fixed units (Caro, Möller et al. 2010).

DES models are technically processed similarly to other individual simulation models (see Figure 6 and Figure 7). To represent variability in the experiences of individuals DES models use random numbers to indicate the expected time of events, resource use and other variable elements. Similarly to other model types, they provide cost and benefits accrued over time; all individuals and events are traceable and as in Markov simulation models, the outputs are aggregated in mean values and distributions of the aggregated values. The outputs of DES can also be expressed in system performance indicators such as resource utilization, wait times and number of entities in lines.

8 Using the term 'entities' instead of 'patients' here, intentionally reflects the much wider range of possibilities provided by DES models compared to state-transition modelling in terms of their application in health care.

**FIGURE 6 REPRESENTATION OF POSSIBLE PATIENT PATHWAYS IN A DES MODEL ABOUT PATIENTS WITH BREAST CANCER**

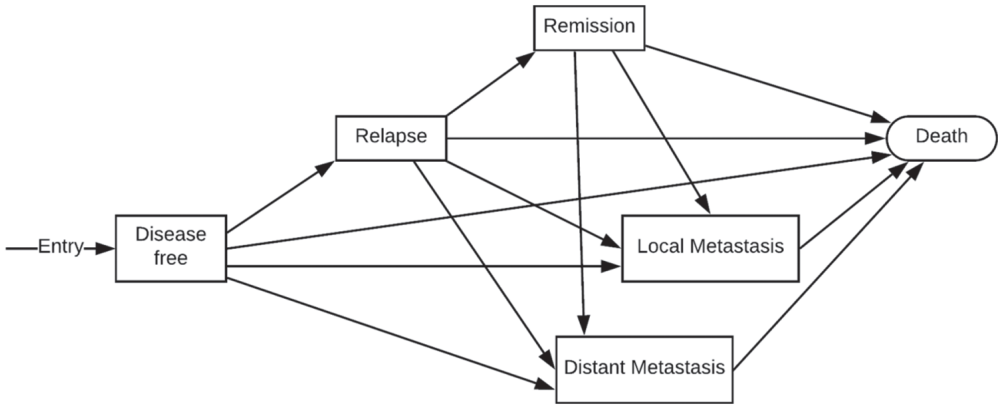
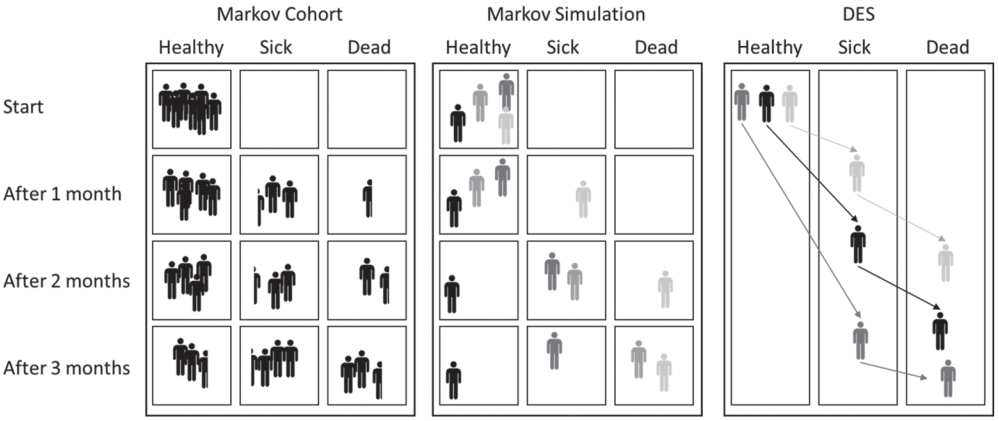



**FIGURE 7 OVERVIEW OF THE DIFFERENT TIME-BASED MODELLING APPROACHES: A) MARKOV COHORT B) INDIVIDUAL LEVEL MARKOV C) DISCRETE EVENT SIMULATION**




Source: adapted from Heeg et al. (2008)

Discrete event simulation is useful for problems in which it is particularly relevant to capture the changing attributes of entities, and in which the processes to be characterized can be described by events rather than health states. DES models can provide enhanced modelling power in applications where exact timing is important while events are quite rare or unpredictable (e.g., a patient might not face an event for 2 years and then a myocardial infarction occurs, with ambulance, treatment, stroke, and other events springing up within a couple of minutes). DES entities in healthcare are usually individual interacting

patients, but these models can also analyze healthcare service system resources, such as doctors, nurses, and ambulances for transport. With regards to the 'queuing' feature two categories of models are distinguished (Karnon, Stahl et al. 2012):

- Non-constrained resource models: they accord with the common structural assumption that all required resources are available as needed, with no capacity limitations. These models are uncommon in non-healthcare applications.
- Constrained-resource models: incorporate capacity limitations. Represent indirect interactions between individuals, generally involving multiple entities competing for access to resources and waiting in queues.

DES is probably the most flexible of all modelling techniques in healthcare decision analysis. It provides a flexible framework to analyze a wide variety of problems. In scenarios where patients' demand for particular resources and their priority status in a queue may be influenced by their attributes, DES is clearly the best choice. DES can also be used to model complex, direct interactions between individuals (e.g., transmission of the disease). While constrained resources pose no problems for most DES tools, special care may be required to model infection dynamics or multiple, correlated health risks (Brennan, Chick et al. 2006) (see section 3.6).

DES models have similar shortcomings to other types of individual simulation models (see section 3.4). An informative DES model requires a significantly richer data source than a typical Markov cohort model. Getting to more details such as moving from state-to-event transition methods may require a greater number of calculations and interactions. Also, accurate representation requires a large enough number of simulation runs to reflect the true variability (the more the variability, the higher the number of runs). Since DES facilitates the representation of complex systems, there is a range of issues along the lines of development modeling, parameter estimation, implementation, analysis, and reporting that should be addressed. The problem of unfamiliarity with DES modelling techniques also implicates a reluctance of analysts to step out of the comfort zone of current modelling techniques (Caro, Möller et al. 2010).

DES models with their substantially increased analytic inputs are definitely not favored when simpler modelling techniques are still appropriate. If describing the "average patient" and the "average treatment effect" without the need to explain correlations, multiple individual characteristics and their relation to risk and treatment effect is sufficient, DES is less preferred. Nevertheless for complex cases properly designed DES provides more accurate and relevant estimates than most modelling techniques.

## 3.6 Dynamic models

The model types discussed so far all provided a portrait of the system at a specific point in time. This so-called *static* approach means that parameters on the system level are assumed to remain unchanged; time does not have an influence on the variables of the system. Static models typically focus on a cohort, either as a whole or as the sum of individuals, that ages as people progress through the model (see sections 3.1–3.5). Any change in the cohort as a whole has no impact on the model variables or on the modelled individuals.[9]

There are areas in healthcare, especially concerning infectious diseases, where the dynamics of the system can have a strong influence on the outcomes of the analysis. When modelling the spread of communicable diseases, analyses often need to reflect the epidemiological effects of the interventions: the rate of infection often determines the number of infected individuals (i.e. the progression of the disease in the community). In such cases individuals who are not reached by an intervention (e.g. vaccination program) can still benefit by experiencing lower risk of infection. This 'herd immunity', defined as the protection for individuals who are not immune due to having a large percentage of a population immune, is a key notion in such models.

Let's illustrate this with the problem of modelling the spread of infectious diseases. On the population level the spread of the disease can be characterized by four distinct phases (Brisson and Edmunds 2003, Briggs, Claxton et al. 2006):

i) pre-vaccination period,
ii) honeymoon, shortly after the vaccination program when the number of infected cases is very low,
iii) post-honeymoon epidemic, when the number of susceptible increases (through births) above threshold that increases the rate of infection to epidemic level,
iv) post-vaccination endemic equilibrium, where long-term equilibrium is reached with lower infection levels than prior to the vaccination program.

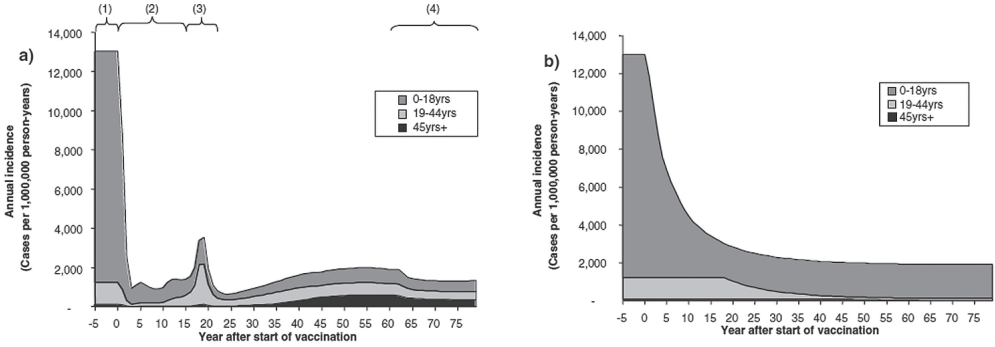
Static models inaccurately capture these phases. They can only assume that the rate of infection among any susceptible population is fixed, and hence they simulate the distinct periods of disease spread. Dynamic models have the feature that the risk of infection is dependent on the number of infectious agents at a given point in time. Population dynamics are affected both by the speed of disease spread and the number of newcomers

9 Note that in most static models the passing of time does have an influence on the cohort's/patients' progression (e.g. age dependent mortality rate). What is meant here is that in dynamic models fundamental (and consequently fixed) characteristics of the cohort change by the passing of time, e.g. the age dependent mortality rate for two identical cohorts is different in a dynamic model because the two cohorts start the model in different years (e.g. the second cohort starts 5 years later) when the system environment has changed (e.g. due to more innovative technologies age-specific mortality rates decrease). Hence fundamental parameters are altered in the dynamic model.

(e.g. birth) – both may change with time. In this way dynamic models allow for the effect of herd immunity and are run over many years on the basis of multiple cohorts.

Figure 8 provides an example for varicella infection control, modelled by using a static and a dynamic model. The various epidemiological phases of disease spread, as pre-vaccination (1), honeymoon (2), post-honeymoon (3), equilibrium (4), are shown for the dynamic (a) and static (b) models. The estimated age-specific incidence of natural varicella after the introduction of infant vaccination (from origo in Figure 8) greatly differs between the 2 modelling approaches: the static model overestimates the incidence of the infection.

**FIGURE 8 PREVACCINATION AND POST-VACCINATION DYNAMICS OF VARICELLA INFECTION USING DYNAMIC (A) AND STATIC (B) MODELLING APPROACHES.**




Key: 1) pre-vaccination period, 2) honeymoon period, 3) post-honeymoon period, 4) equilibrium.

Source: Brisson et al. (2003)

Dynamic modelling is not necessarily a must when modelling all infectious diseases, or similar types of situations. Static models are still acceptable if target groups eligible for the intervention are not epidemiologically important (e.g., evaluation of hepatitis A vaccination in travelers from low- to high-incidence countries), or when effects of immunizing a given group are expected to be almost entirely direct (e.g., vaccination of the elderly against influenza or pneumococcal disease) (Pitman, Fisman et al. 2012). But where static models project interventions to be unattractive or borderline attractive (i.e., close to willingness-to-pay thresholds) supplementary dynamic modeling is often recommended as an alternative to evaluate whether the inclusion of time dependent system-level variables alter the projected outcomes (Pitman, Fisman et al. 2012).

It is important to note that, technically the execution of dynamic models can take many forms. They may be performed with (Pitman, Fisman et al. 2012):

- deterministic or stochastic results,
- individual or cohort-based simulation,
- economic, health or standalone epidemiological outcomes,
- simple explorations of the system or a very detailed analysis with many parameters.

Dynamic modelling is not restricted to decision modelling of healthcare interventions. It has great potential in further areas of healthcare, such as modelling physiological interactions in the body that affect treatment outcomes or networks of related diseases (Brennan, Chick et al. 2006). They can also be used to examine the evolution of complex systems, processes and interactions between entities. Because of the time component, dynamic models can provide a representation of the evolution systems and this generally allows for more accurate predictive properties.

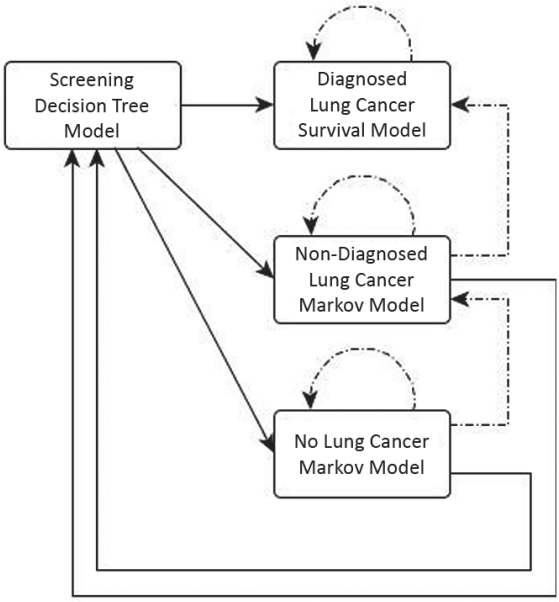
## 3.7 Combining/Hybrid models

In this chapter distinct forms of decision analytic models have been discussed so far. Nonetheless, rather than using a single type of modelling approach to describe a decision problem, a combination of these techniques is often more viable; especially when the decision problem necessitates models which are built up using the combination of multiple modelling techniques.

Figure 9 provides an example of combining various modelling techniques. The analysis of population screening with low dose CT to detect lung cancer is carried out via modelling a one-time procedure and the evolution of a progressive disease. First, patients enter a decision tree model in which they face screening for lung cancer with low dose CT followed by other confirmatory diagnostic processes. Then patients based on the diagnosis are followed in various sub-models: i) Markov model for patients without lung cancer ii) Markov model for patients with undiagnosed lung cancer iii) cancer stage-based survival model for patients with diagnosed lung cancer. Patients can be processed through the model as a cohort or as individuals; the latter technique provides more flexibility to take into account the heterogeneity (subgroups) of patients and their patient history tracks. A similar approach for the case of diabetes screening and for modelling long-term ADHD is provided in Appendix IV.

As it will be discussed in the next chapters, choices on appropriate techniques and their implementation depends on a number of circumstances and successive decisions.

**FIGURE 9 THE MODEL OF LOW DOSE CT SCREENING COMBINING A DECISION TREE, A SURVIVAL, AND TWO MARKOV MODELS**




Source: based on Vokó et al. (2017)

# 4 Building decision models

Balázs Nagy and Ahmad Fasseeh

Building decision models requires careful preparation with regards to setting up the model concept, planning and coordination of model development, analysis of outcomes and validation of the results. Methods are strongly determined by the research question, context and resources. At the same time there are common rules to be followed by model developers.

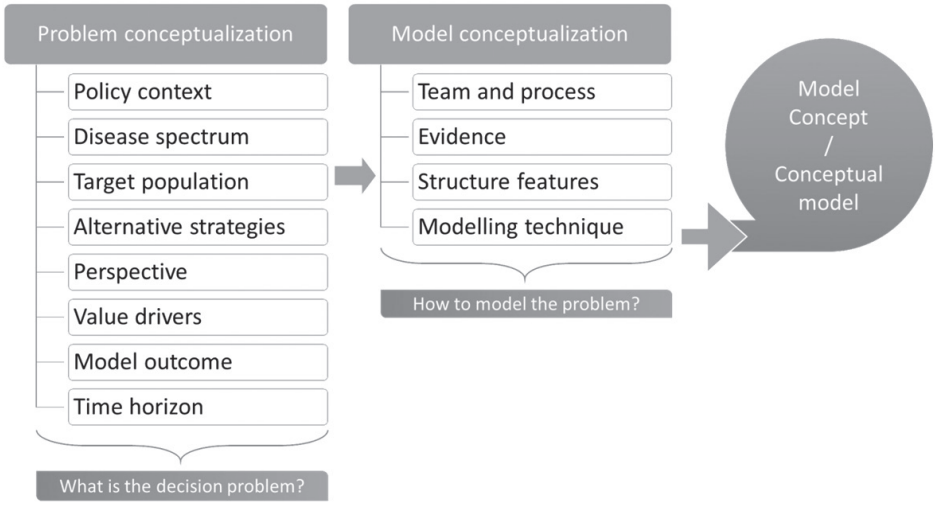
## 4.1 The model concept

The development of good decision models starts with conceptual modelling as a first step. This essentially requires the developer to understand the complexity of the 'real-world' that will have to be represented. Then choices available for translating this 'world' into a credible conceptual and mathematical structure need to be explored (Tappenden 2012). As a result of these steps the model is abstracted from a real or proposed system with simplification and assumptions – based on what is not known about the real system (Robinson 2008). The essence of good conceptual modelling is to get the level of simplification correct, i.e. the modeler has to abstract at the right level (Robinson 2010).

The model concept is fundamentally a theoretical construct, representing (often visually) the processes, relationships, and variables considered to be important within the system under scrutiny. It describes without technical specification the objectives, inputs, outputs, content, assumptions and simplifications of the model. The concept development is both driven by needs and conditions, and it both drives and is driven by the variables that are considered important in the world to be abstracted (Group 2010).

Two phases of model concept development is distinguished by Tappenden (Tappenden 2012): problem conceptualization and model conceptualization (Figure 10).

**FIGURE 10 KEY DETERMINANTS OF CONCEPTUAL MODEL DEVELOPMENT**




*Note:* This Figure synthesizes two sources: Tappenden 2012, Roberts and Russel et al. 2012

During the **problem conceptualization** the modeler, in conjunction with other stakeholders, determines what is relevant to the decision problem, and at the same time, what can reasonably be considered irrelevant. This process builds upon several factors (adapted from Roberts, Russell et al. 2012):

- *Policy context*: the model scope and the structure should be consistent and adequate with the decision problem and its environment – including the funder, the developer, the policy audience and whether the model is for single or multiple applications.
- *Disease spectrum*: the model should represent disease processes appropriately; it should address all disease processes which are necessary to characterize the specific healthcare program of interest.
- *Target population*: the model population should be defined in terms of features relevant to the decision such as geography, patient characteristics, including comorbid conditions, disease prevalence and disease stages.
- *Alternative strategies and interventions*: choices on the comparators of the healthcare program should be driven by the nature of the problem, not solely by data availability or quality. All feasible and practical strategies should be considered.
- *Perspective of the analysis*: model outcomes should be consistent with the perspective stated and defined. Included and excluded outcomes should be determined in relation to the perspective.
- *Value drivers*: crucial features of the assessed technologies having influential impact on the model outcomes should be taken into account without exception. As

a general approach, there is a value in having any sort of underlying biological or clinical process with significant influence on the model outcomes.

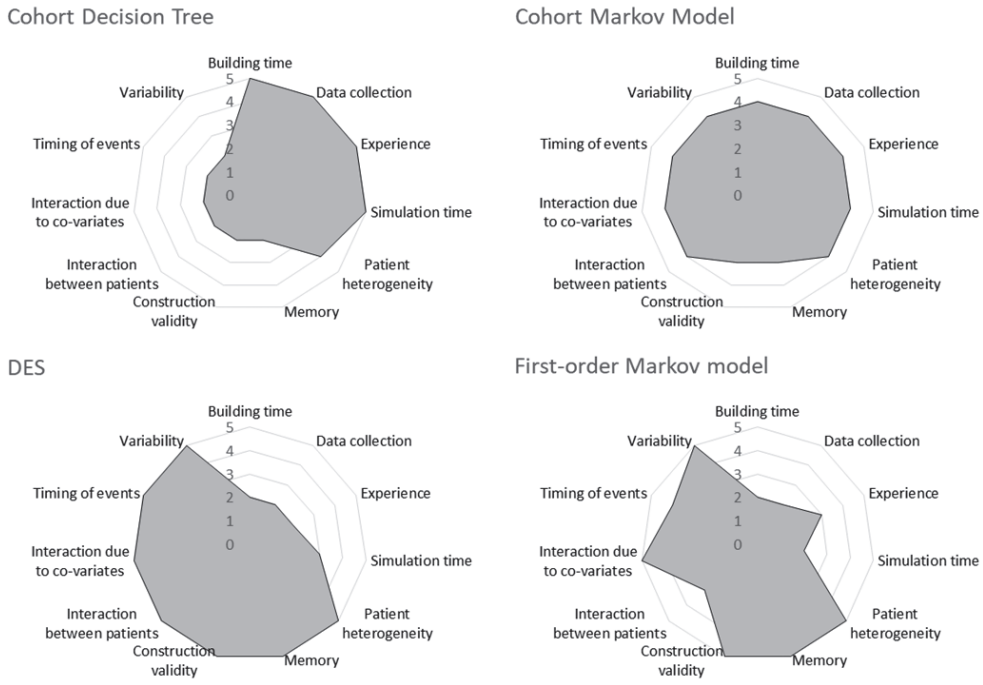
- *Model outcomes*:
  - health outcomes may be events, cases of disease, deaths, life-years gained, quality-adjusted life-years, disability-adjusted life years, or other measures important to stakeholders, should be directly relevant to the question being asked.
  - resource use and costs of interventions in the analysis should be clearly defined in terms of frequency, component services, dose or intensity and duration.
- *Time horizon of the analysis*: this should be long enough to capture relevant differences in outcomes across strategies. It should be set as long as any change in the difference between the outcomes of the competing strategies is observed.

**Model conceptualization** represents the components of the problem by presenting particular analytic methods and processes and directs the decision as to which modeling technique to use. Crucial stages of model conceptualization are as follows (Roberts, Russell et al. 2012, Tappenden 2012):

- *Set up the team and process*: work can relate to expert consultations, influence diagrams, concept mapping, or any other method which converts the problem conceptualization into an appropriate model structure, ensuring it reflects current disease knowledge and the process modeled. All resources (personnel, time, software etc.) to carry out the work should be identified.
- *Review the evidence*: any decision should carefully consider the possible sources of evidence to inform conceptual models. These sources include:
  - clinician inputs
  - existing systematic reviews
  - clinical guidelines
  - existing efficacy studies
  - existing economic evaluations or models, and
  - routine monitoring sources.
- *Specify structure features* based on several factors such as
  - unit of representation: individuals or groups,
  - interactions between individuals,
  - time horizon and time measurement,
  - resource constraints (if any).
- *Determine modelling technique*: for some problems certain types of models, such as decision trees or Markov models, for other problems, combinations of model types, hybrid models and other modeling methodologies might be appropriate. Such judgements are based on series of conditions which are often trading off against each other (see more in 4.2). An example of such trade-offs is seen on Figure 11 which compares the strengths and weaknesses of using decision tree, Markov

cohort, Markov simulation, and discrete event simulation techniques. The further away from the center the relevant line covers the specific axis, the better this kind of modelling is compared to the others with respect to the particular modelling characteristic. Important to highlight that these criteria do not necessarily have the same weight and one may depend on another.

**FIGURE 11 STRENGTHS AND WEAKNESSES OF COHORT DECISION TREES, COHORT MARKOV MODELS, MARKOV SIMULATION MODELS AND DISCRETE EVENT SIMULATION MODELS FOR APPLICATION IN A CHRONIC COMPLEX DISEASE SUCH AS SCHIZOPHRENIA.**




Source: adapted from Heeg et al. (2008)

Selecting the appropriate level of detail is one of the most difficult decisions developers face. The model must be complex enough to capture the differences in value (e.g. health gains or cost savings) across the compared strategies and provide the ability to cover all important dimensions of reality to make right decisions. Models that are too complex may be difficult to build, debug, analyze, understand, and communicate. Simplicity is also desirable for transparency, ease of analysis, validation and description. However, simplicity cannot overrule the aspiration for an adequate level of accuracy. Models that are too simple may lose face validity because they do not incorporate all aspects recommended by

content experts (e.g. by clinical experts and other healthcare professionals). In particular, if the experts and practitioners within the system do not trust the conceptual model, it will remain unused, regardless of its quality.

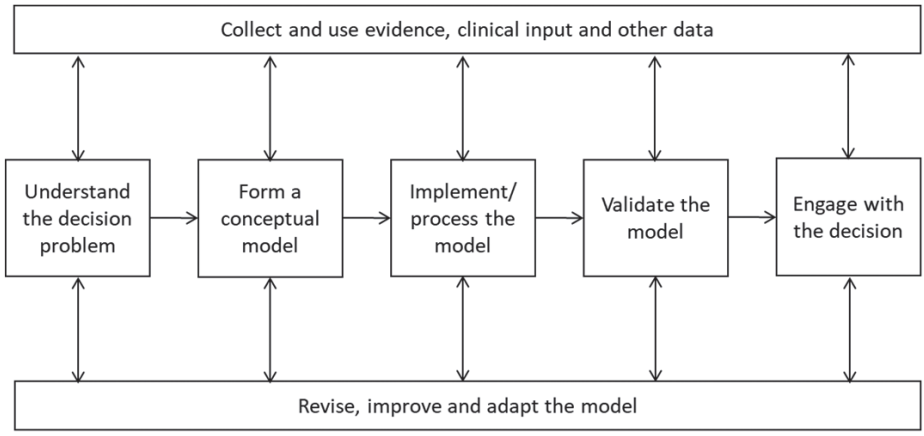
Model building is a fundamentally iterative process. So, whether conceptual modelling is being performed formally or informally, it continues to be refined before, during and after the model has been developed (Robinson 2010) (see more in section 4.2). Moreover, it may happen that while building a model based on an initial concept, it becomes clear that the chosen approach is not appropriate and a new modelling approach has to be chosen to transform the concept into a usable model (Group 2010).

## 4.2 Model development process

Despite of the specifics of process and methods, it is difficult to set out distinct methods and procedures which will unquestionably have to be followed. The 'art' of model building is largely learned by experience and strongly depends on the ability of the modeler to synthetize the various aspects of model development into an optimal solution.

The process of model development[10] is an iterative exercise that does not linearly proceed with time, tasks and other working blocks. The process starts at the point of thinking about the analysis and is not finalized perhaps until quite close to the end of the whole project. Modelers constantly need to seek for advice about whether or not they have proceeded the correct way and used the best available methods to replicate reality. This is profoundly iterative process has a number of crossroads, junctions and possible U-turns. Steps of model development are shown on Figure 12.

**FIGURE 12 THE STEPS OF THE MODEL DEVELOPMENT PROCESS**




10 Including model conceptualization, as seen in Chapter 4.1

### 4.2.1 Understanding the decision problem

Understanding the decision problem (also called problem conceptualization, see 4.1) is part of developing the model concept. The decision problem can be presented as a construct representing (often visually) the processes, relationships and variables considered to be important. The scope of a modelling problem helps define the boundary and depth of a model and identifies the critical factors that need to be incorporated. Understanding the decision problem may involve several activities (Chilcott, Tappenden et al. 2010):

- setting up the research question,
- engagement with clinicians,
- engagement with decision-makers,
- engagement with methodologists (e.g. analysts, modelers),
- gaining an understanding of what is sufficient and feasible.

Models intended as "multipurpose" tools that start without a clearly defined question generally end up without any clear conclusions (Chilcott, Tappenden et al. 2010). Conversely, models designed with a clear purpose in mind, once validated, can often be easily adapted to other purposes (Group 2010). More details about problem conceptualization were shown in section 4.1.

### 4.2.2 Forming the conceptual model

Once the decision problem has been identified the information surrounding the decision needs to be synthesized and converted into a particular analytic method. This is expected to provide the technical framework of the analysis (see also model conceptualization in section 4.1). This process considers the available inputs and the applicable design together and in relation with the decision problem.

Selecting the appropriate modelling technique is driven by many factors (see all in 4.1). Very few research questions have a unique solution and most healthcare decision problems can be solved in more than one way. Likewise, more than one modelling approach is possible, each having advantages and disadvantages. Similar results might be obtained using different modelling techniques which is a good indication of making the right choices on the concept and structure (see more in chapter 0).

### 4.2.3 Processing and implementing the model

Following the model concept the specified model has to be implemented in such a way as to produce answers (e.g. predictions) with regards to the questions of interest. The idea is that everything on paper has to

i) be translated to a mathematical construct,
ii) be populated with data and
iii) provide sensible and interpretable results.

While setting up the model concept and choosing the appropriate modelling structure and technique are rather straightforward processes, things can become less foreseeable when it comes to implementation. From this point on, the number of iterations between the various model building blocks (see Figure 12) significantly increases. There is a lot of interaction, especially with the blocks of data collection and validation. The process of populating a model involves bringing together all relevant evidence and synthesizing them appropriately given the modelling framework and parameters. A populated model helps determine which variables are important to characterize the decision problem, test decision validity, and tune the model to make appropriate predictions.

The required level of mathematical and computational programming is strongly driven by the complexity of the model, the programming language and the software environment in which the calculations are embedded. There are dozens of software environments in which healthcare models are developed. In most cases Microsoft Excel using VBA programming is sufficient. However, there are other, user friendly modelling solutions, such as Treeage, ARENA, heRo3 and Simul8, applicable for healthcare which in certain situations might be more efficient than the traditional Excel-based programming. When necessary, modelers may choose platforms requiring the use of programming languages like the very flexible R mathematical package or programming languages as Java, Python or C++.

### 4.2.4 Validate the model

During and after the implementation of a model it has to be tested for validity. This process substantiates that the model performs with satisfactory accuracy within the domain of its applicability. This includes engaging with clinical experts to check face validity, testing extreme values, checking logic, data sources etc. The validation process involves several methods and is carried out (by applying different elements and methods) throughout the entirety of the model building process. See more details on model validation in section 0.

### 4.2.5 Engaging with the decision

Once the finalized model is implemented, tested and tuned it can be applied to support the decision-making process. This phase mostly concerns the reporting and use of the model bearing in mind the decision-making rules. Outcomes have to be able to answer specific questions and the analysis has to provide details about the uncertainty around the model estimates as well.

Nevertheless, the model does not only provide explicit results, but it should provide convenient tools to explore certain properties of system behavior as well. For example, if a healthcare ministry has a budget for only a fixed number of physicians, they may wish to know where to locate those physicians in order to achieve optimal patient care. In general, the process of model optimization is applied to answer this type of question. Often the question can be written as "which selection of parameters minimizes the cost such that the

desired result occurs?" Finding the answers to such questions has become a field in itself and can be accomplished by a number of different means (Group 2010).

The most common method to explore uncertainty and model behavior is via sensitivity and scenario analyses (see more in section 5.2). It is often useful to test parameters with extreme values to see if they change the conclusion of the analysis (i.e. change the decision).

Finally, no matter how well a model is tested, tuned, and implemented, it can only examine the aspects of the system it was designed to study. After a model is created and final results are obtained, a common mistake is to over-interpret the importance of the results and assume causality where only association is present (Group 2010).

### 4.2.6. Collect and use evidence, clinical input and other data

It is often said that models are only as good as the data used to test and tune them. Data collection is a strong influential factor in healthcare model development and the quality of data and subsequent analysis often imposes limitations on the quality of models and their results. As a rule of thumb, models are designed and tested with all data that are available; the final implementation uses the most feasible and appropriate set of variables.

Information on model variables is often extracted from data collected for other purposes; as a result, data may be biased, inaccurate or contain errors. The modelling framework should allow for the handling such discrepancies through integrating tools by which the most precise estimation can be achieved. Several models for example use competing methods to estimate patients' quality of life, by either using multiplicative or additive aggregation techniques. Supporting tools often come from methods of evidence synthesis such as meta-analysis, network meta-analysis, mixed treatment comparison, utility mapping and Bayesian statistics.

Models in healthcare are often built on a set of assumptions, some of which are testable and others that are not. These assumptions must be clearly stated and, whenever possible, tested. Often supporting statistical analysis will inform the modeler that some of their basic assumptions about the system were wrong, forcing the modeler to take a step backwards and form a new conceptual model for the problem. This may occur when a modeler determines that a variable assumed to be insignificant turns out to be significant or vice versa.

It is not difficult to conclude that generating evidence should be carried out throughout the entire process of modelling and can be influential in all steps, as illustrated on Figure 12.

### 4.2.7 Revise, improve and adapt the model

Regardless of the computational environment, the data being used and the assumptions made, the modeler will need continuous feedback on the development process. Are the formulas correct? Do the input data and assumptions reflect reality? Are the results interpretable, and do they make sense to support the purpose of the analysis?

As discussed earlier even after the first model results the model structure, methods or some of the inputs and assumptions might be reconsidered and minor or even major changes are initiated. Such activities are often commenced after the core model is completed and the first adaptation to another environment is carried out.

Another driving factor of model improvement is directed by the raising of further research questions. The efforts to answer these novel questions by testing the model in new circumstances could help find logical or input discrepancies.

# 5 Handling uncertainty in decision models

Balázs Nagy, Ahmad Fasseeh and László Szilberhorn

Chapter 1 discussed that choices have to be made by healthcare decision-makers now and not later. There should be a clear 'yes' or 'no' answer to reimbursement and other resource allocation questions. In order to make the best decision one needs to be aware of the level of confidence in the findings.

All models are limited in their capacity to capture real-life circumstances. This may happen due to lack of good quality evidence, inappropriateness of the data, uncertainty around long-term prediction, or for other reasons. While in reality, the value of a model input parameter (e.g. cost of hospitalization) is distributed around a mean the modeler's analysis is often based on single point estimates (such as the mean or median value). Such 'deterministic' values do not inform us about the variability in data and sources. Ambiguity around the model structure, the extrapolation methods or other factors can also be a subject of uncertainty. In the end deterministic results are able to reflect only a portion of uncertainty around selected components of the analysis (Berger, Bingefors et al. 2003). Hence, economic models are strongly advised to assess uncertainty around all possible aspects of the analysis.

Methods for showing and handling uncertainty are collectively referred to as a sensitivity analysis. Such analyses serve two main purposes: i) assess the confidence in the chosen course of action supported by the model and ii) ascertain the value of collecting additional information to better inform the decision (Briggs, Weinstein et al. 2012). These are discussed in the following sections.

## 5.1 What is uncertainty?

In the context of decision analytic modelling, uncertainty can be defined by using four broad categories (Briggs, Sculpher et al. 1994, Gray, Clarke et al. 2010, Briggs, Weinstein et al. 2012):

– variability (or stochastic uncertainty),

– heterogeneity,
– parameter uncertainty and
– model (structural and methodological) uncertainty.

The health economic literature has discussed these four broad terms in a variety of ways. In the context of decision analytic modelling, uncertainty is mostly referred to as having imperfect information about the precise values of the parameters of interest, or about the applicability of methods being used to design and build the model. This definition restricts the focus to parameter and model uncertainty and omits variability and heterogeneity.

*Variability*, in this context, refers to the inherent random variation between different subjects (see more in section 5.1.1) and *heterogeneity* relates to variation between subjects that can be explained and attributed to specific factors (see more in section 5.1.2). These two categories are important to understand uncertainty, however, in the context of economic evaluations, they do not necessarily contribute to the uncertainty defined above. Nevertheless, we still consider these categories, to the extent it is deemed important to understand the broad picture of uncertainty for economic modelling in healthcare.

The four concepts of uncertainty are summarized in Table 3 and in the forthcoming paragraphs. Also analogous concepts used in the field of regression analysis are presented in Table 3 to support explanation.

**TABLE 3 TYPES OF UNCERTAINTY FOR DECISION MODELLING: CONCEPTS, SYNONYMS AND ANALOGIES**

| term | concept | synonym term sometimes employed | analogous concept in regression analysis |
|---|---|---|---|
| **Stochastic uncertainty** | random variability in outcomes between identical patients | variability, Monte Carlo error, first order uncertainty | error term |
| **Parameter uncertainty** | the uncertainty in estimation of the parameter of interest | second order uncertainty | standard error of the estimate |
| **Heterogeneity** | variability between patients that can be attributed to characteristics of those patients | variability, observed or explained heterogeneity | beta coefficient (or the extent to which the dependent variable varies by patient characteristics) |
| **Structural uncertainty** | the assumptions inherent in the decision model | model uncertainty | the form of the regression model (e.g. linear, log linear) |

Source: adapted from Briggs et al. (2012)

### 5.1.1 Variability, stochastic uncertainty

We can think about variability as the variation or randomness we observe when recording information on resource use or outcomes within a homogenous sample of patients (Groot Koerkamp, Weinstein et al. 2010). This uncertainty is entirely due to chance. It is also referred to as first order uncertainty or stochastic uncertainty (Briggs, Weinstein et al. 2012). It relates to the fact that individuals facing the same probabilities and outcomes will experience the effects of a disease or intervention differently[11]. This type of uncertainty reflects the inherent variability that exists in the parameters of interest between patients within a population (Andronis, Barton et al. 2009). Consequently, variability is reflected in the standard deviations associated with the mean value. In regression analysis this type of uncertainty is analogous to the error term (see Table 3). If a stakeholder is not interested in individual patient outcomes, then the analysis of stochastic uncertainty in patient level models only increases 'noise' around the expected outcomes (Groot Koerkamp, Weinstein et al. 2010).

### 5.1.2 Heterogeneity

Heterogeneity relates to observed differences between patients which can in part be explained by their characteristics (e.g., age- and sex-specific mortality) (Gray, Clarke et al. 2010, Briggs, Weinstein et al. 2012). For example, the hospital costs arising as a result of myocardial infarction may differ between young and old patients because older patients typically spend a longer time in the hospital. There will still be variability between patients within each of these subgroups in terms of whether or not they will experience a particular outcome over time. However there is no uncertainty here considering that the baseline characteristics will be known with certainty. Baseline patient characteristics can influence each estimated parameter in the model: for example, we can distinguish heterogeneity in treatment effects, in costs, and in utilities (Groot Koerkamp, Weinstein et al. 2010).

As was pointed out in the introduction of this section, heterogeneity is not a source of uncertainty as it relates to differences that can in principle be explained. Its relevance lies in the identification of subgroups for whom separate cost-effectiveness analyses should be undertaken. Such analyses may inform us of alternative decisions regarding the service provision to each subgroup, or may contribute to a weighted analysis of the aggregate group (Briggs, Weinstein et al. 2012). The analogous term in regression analysis is the beta coefficient or the extent to which the dependent variable varies according to patient characteristics (see Table 3).

11 Just as a fair coin might come up heads or tails on any given toss (e.g., the first patient in a sample might respond to a treatment but the next may not; the first may not experience an adverse effect but the second may; the first may stay in hospital for 2 days and the second for 3 days)(Briggs 2012).

### 5.1.3 Parameter uncertainty

Parameter uncertainty relates to the accuracy in the parameter/input estimates that is used for the analysis, such as transition and event probabilities, costs, utilities, treatment effects, or the mean length of hospital stay. It reflects the uncertainty that arises from the imperfect knowledge of true values of the parameters that are used in the analysis (Andronis, Barton et al. 2009). The term parameter uncertainty is not equal to the uncertainty around the realization of individual events or outcomes (see first order uncertainty in section 5.1.1). Parameter uncertainty is also referred to as second order uncertainty[12] (Briggs, Claxton et al. 2006). It is analogous to the term standard error in regression analysis which calls to mind the distinction between standard deviation (estimate of how individual observations within a sample vary) and standard error (precision of an estimated quantity) (Briggs, Weinstein et al. 2012).

There is no doubt that this type of uncertainty needs to be reflected in the cost-effectiveness analysis. This can either be represented via a deterministic sensitivity analysis (DSA) or a probabilistic sensitivity analysis (PSA) (see more in sections 5.2.1 and 5.2.2). The latter facilitates the estimation of additional uncertainty measures such as the expected value of perfect information (EVPI), which may be estimated for the model as a whole, or for specific parameters or sets of parameters (expected value of partial perfect information, see more in section 5.3).

### 5.1.4 Model uncertainty

The model, once finished, is still dependent on a number of assumptions. Decisions have to be made about the natural course of the disease, the impact of medical interventions, the clinically and economically meaningful outcomes, and the inclusion of relevant events, comparators and use of statistical estimation methods. The final form of the model is also influenced by the complexity to be reflected by the analyst and the time available for development (see in section 4.1 and 4.2). Uncertainties around the structure, the methodology and the process all contribute to the term we refer to as *model uncertainty*.

As the subset of model uncertainty, *structural uncertainty* concerns the decisions and assumptions we make about the structure of the model such as the inclusion of relevant states, the established links between states and also the way the intervention and disease pathways are modeled.

Another subset of model uncertainty is *methodological uncertainty*. This concerns the combination of methods used by the analysts who carry out the analysis; i.e. if the analysis was conducted again by another team of analysts, the results might be different due to the use of other techniques/data. This is the case, for example, when patient's survival is

12 Following the example in footnote 11, no one will doubt that a head is a head and a tail a tail (second order uncertainty), but one may doubt what a throw of a fair coin will show (first order uncertainty). If one does in fact doubt the fairness of the coin, then we refer to second order uncertainty.

predicted first by using independent survival curves or alternatively by a Cox proportional hazard model – two fundamentally different methodologies.

The third subset of model uncertainty is *process uncertainty* which relates to questions about the appropriate methodology to combine the input parameters (Andronis, Barton et al. 2009). Such uncertainty occurs when, for example, model states can be designed i) to appropriately correspond with quality-of-life estimates available in the literature or ii) rather to match with important cost items (e.g. surgical interventions). Both approaches will have different consequences regarding the use of further inputs and also concerning the final model estimates.

Model uncertainty can be as important as parameter uncertainty and any disagreement about these features may be a reason to undertake a sensitivity analysis (Briggs, Weinstein et al. 2012). The analogous term to model uncertainty in a regression analysis is the exact form of the regression model (see Table 3).

### 5.1.5 Handling uncertainty in decision models

The impact of uncertainty should be assessed in decision models. The most influential parameters and their impacts on the model results should be identified, quantified, and interpreted. Depending on the type of uncertainty there are various ways to present and deal with uncertainty issues (Briggs, Claxton et al. 2006, Briggs, Weinstein et al. 2012):

- *Variability* in most decision models (especially simulation models) is largely a by-product of the modelling process rather than an attempt to capture a real-world phenomenon and so the analytical concern is to take steps to eliminate variability from the results. This, for example, in individual simulation models can be done by increasing the size of the simulated sample (see in section 3.4).
- *Heterogeneity* should be handled by providing a flexible modelling framework. A model may be rerun for different subgroups, or an overall measure of cost-effectiveness (across the entire population) can be reported together with variability measures (e.g. standard error of the mean) due to patient heterogeneity. Heterogeneity can also be handled by making model parameters a function of other parameters: e.g. basic transition probability is a function of age or disease severity. The process involves simultaneously switching the parameter values and assumptions according to the corresponding subgroup of interest.
- *Parameter uncertainty* is the most widely researched and published area of uncertainty which is handled through either deterministic or probabilistic sensitivity analyses (see more in section 5.2).
- *Model uncertainty* is a moderately researched area, although its influence on the model results can be even greater than for all other types of uncertainty. Model uncertainty is usually handled by performing various types of deterministic sensitivity analyses and scenario analyses (see more in section 5.2).

## 5.2 Sensitivity analysis

In sensitivity analysis (SA) model parameter estimates are varied across a range to determine the impact of their change on the model outputs (Briggs, Weinstein et al. 2012). The process can be carried out on parameter values, assumptions, methods, or anything else which can be varied within a reasonable range. Two major types are distinguished:

- *Deterministic sensitivity analysis (DSA)* evaluates the influence of uncertainty in one or more parameters on the expected outcomes (Groot Koerkamp, Weinstein et al. 2010). These parameters are manually changed usually across a pre-specified range.
- *Probabilistic sensitivity analysis (PSA)* is the stochastic evaluation of the model which permits the joint uncertainty across all parameters to be assessed at the same time. It involves sampling model parameter values from distributions imposed on model variables and the generation of cost and effectiveness estimates (Gray, Clarke et al. 2010).

DSA's and PSA's main focus is on the analysis of parameter uncertainty and they put less emphasis on the analysis of other types of uncertainties (as discussed earlier in section 5.1).

### 5.2.1 Deterministic sensitivity analysis

DSA in its simplest and most commonly used form is carried out via a *one-way (or univariate) deterministic sensitivity analysis*: the value of one variable is varied independently and singly within a plausible range, while the other variables are kept constant. The range of variation of each parameter is usually pre-specified, and where appropriate it corresponds to the uncertainty in that parameter reported in source studies. It can be from highest to lowest if a range of estimates is available or within 95% confidence limits if reported or simply within a plausible range which could be arbitrary (Gray, Clarke et al. 2010).

In a *multi-way (or multivariate) deterministic sensitivity analysis* more than one parameter estimate is varied simultaneously. A two-way analysis is useful to test the two most important parameters from the perspective of the analysis (e.g.: the main value driver of the intervention and price in an early-phase pricing model). In a *scenario analysis* multiple variables are changed to form a distinct alternative case compared to the base case analysis. Parameters can simultaneously be set for extreme scenarios: optimistic *best case scenario*, or pessimistic *worst case scenario*. Another form of sensitivity analysis is *threshold analysis* in which the critical value of parameters changing the decision are identified – e.g. set of parameter values resulting in the support or rejection of a reimbursement decision.

One-way and multi-way sensitivity analyses may be carried out on a sequential basis. A one-way sensitivity analysis can be graphically presented on a *tornado diagram*: the

most critical variables in terms of impact on the model outcome are at the top of the graph and the rest are ranked according to their impact thereafter (see Figure 13). The tornado shape arises by ordering the bars by their width, starting with the widest at the top. In the two-way sensitivity analysis the calculated ICERs can be presented in a *matrix*-like framework where the rows and columns provide the results of changing two variables together (see Table 4).

**FIGURE 13 TORNADO DIAGRAM ILLUSTRATING THE 15 MOST INFLUENTIAL VARIABLES OF A COST-EFFECTIVENESS MODEL**

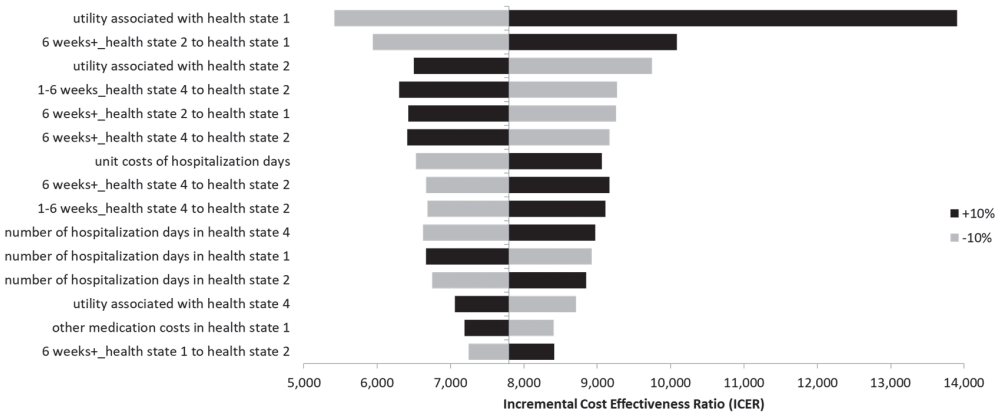



**TABLE 4 RESULT OF THE TWO-WAY SENSITIVITY ANALYSIS IN AN EDUCATIONAL MODEL**

| Drug Price → | $800 | $1,000 | $1,200 | $1,400 | $1,600 | $1,800 | $2,000 |
|---|---|---|---|---|---|---|---|
| Effectiveness ↓ | | | | | | | |
| 80.0% | $436 | $2,512 | $4,587 | $6,663 | $8,739 | $10,814 | $12,890 |
| 77.5% | $3,193 | $4,875 | $6,558 | $8,241 | $9,924 | $11,607 | $13,289 |
| 75.0% | $5,072 | $6,487 | $7,902 | **$9,317** | $10,732 | $12,147 | $13,562 |
| 72.5% | $6,435 | $7,656 | $8,876 | $10,097 | $11,318 | $12,538 | $13,759 |
| 70.0% | $7,469 | $8,542 | $9,616 | $10,689 | $11,762 | $12,836 | $13,909 |

For a deterministic sensitivity analysis a clear and full justification for the choice of variables is required. Also a clear explanation of the information source used to specify the ranges is necessary. When the sensitivity analysis involves an analysis of extremes, the analysts should justify the extreme values chosen and provide a clear presentation of the analysis in order to allow the reader to assess the analysis relative to their own context. When the value of a model parameter is indeterminate, a threshold analysis is particularly

useful, but there is a need to provide a clear rationale for, and definition of, the threshold applied (Andronis, Barton et al. 2009).

### 5.2.2 Probabilistic sensitivity analysis

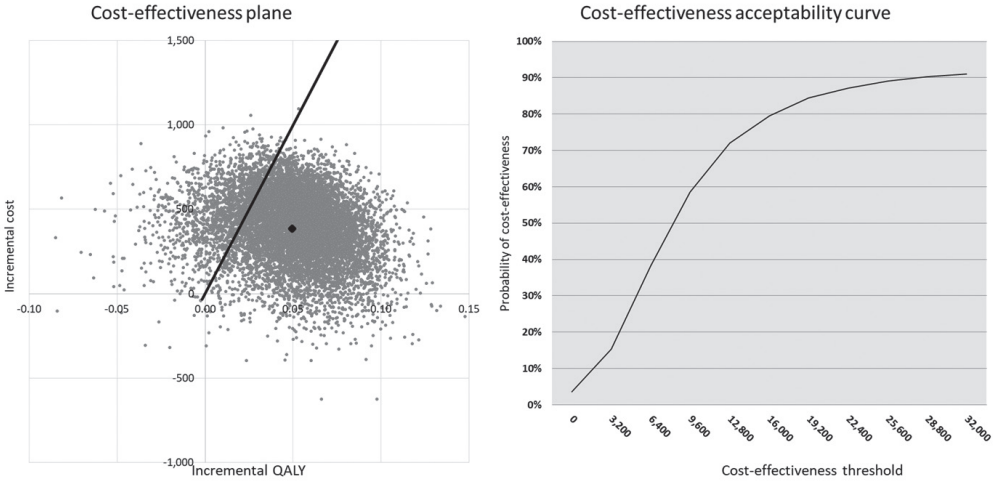
A more complete assessment of parameter uncertainty in health economic models calls for a probabilistic sensitivity analysis. PSA assigns a specified distribution to each input parameter and, by drawing randomly from those distributions, generates a large number of mean cost and effectiveness estimates that can be used to form an empirical joint distribution of the differences in cost and effectiveness between interventions (Andronis, Barton et al. 2009).

The process starts by specifying a probability distribution for each parameter of interest. Each distribution represents both the range of values that the parameter can take, as well as the probability that it takes any particular value. Then by running a so-called second order Monte Carlo simulation, a value is selected for each parameter from its individual probability distribution. The analysis is repeated a large number of times to propagate uncertainty and present a distribution of possible payoffs associated with the technologies of interest. In this way PSA will present the realization of the uncertainty that exists in the analysis as characterized by the probability distributions.

To run the PSA appropriately, evidence-informed distributions should be placed around all uncertain model parameters while any excluded parameters must be justified. The distributional assumption for each variable should reflect the nature of the variable; i.e. it should be consistent with any logical bounds on parameter values given its nature (e.g. utility scores with upper bound of 1, costs >= 0). When correlation between variables is expected, joint distributions should be used and independence should not be assumed (Andronis, Barton et al. 2009). There are often rules defined in methodological guidelines on choosing the appropriate distributions for different types of parameters. Appealing to the Central Limit Theorem, an appropriate distribution for any parameter includes the normal distribution, but given constraints on logical bounds for certain parameters, other distributions may be a better choice.

The result of the PSA is most commonly scatter plotted on a so-called cost-effectiveness plane (see Figure 14). The plots on the plane present all the results for the outcomes of interest (usually incremental costs and incremental quality adjusted life years [QALYs]) with respect to the compared technologies. The scatter plotted results can also be summarized according to their relation to the willingness-to-pay threshold (e.g. how much the society is willing to pay for one additional QALY). This can be illustrated on the cost-effectiveness acceptability curve (CEAC) (Figure 14). The CEAC plots the probability that one treatment is cost-effective compared to another as a function of the willingness-to-pay threshold for one additional unit of efficacy (Berger, Bingefors et al. 2003). The CEAC is in many ways the most helpful expression of the relative cost-effectiveness comparison between competing treatments (Briggs, Weinstein et al. 2012).

**FIGURE 14 COST-EFFECTIVENESS PLANE AND COST-EFFECTIVENESS ACCEPTABILITY CURVE FROM A COST-EFFECTIVENESS MODEL**




## 5.2.3 Application of DSA and PSA

DSA can give insight into the factors influencing the results and can also provide a face-validity check to assess what happens in case of changing inputs or assumptions. Where the direction and magnitude of change in outcome tied to the change in each model parameter are reasonable and justifiable, there is a good chance of having no systematic error in the model. Tornado diagrams and other tools help decision-makers be explicit with respect to the key drivers of uncertainty in the model and provide a simple way to summarize and depict the impacts of different variables underlying an analysis. It presents a useful tool to summarize and portray the uncertainty and provides an initial, semi-qualitative assessment of uncertainty. It provides a natural starting point for the investigation of uncertainty and provides a standard route through which some of the key drivers of the cost-effectiveness results should best be revealed. It is a useful tool to identify critical model parameters and as a matter of fact, it is inevitable and mandatory for the full analysis of cost-effectiveness models (Gray, Clarke et al. 2010).

DSA is imperfect in several ways, however. It can only represent the impact of change in certain predetermined directions. Reality is typically more complex: many different combinations of variables may be possible and variables may be associated with each other. By its nature DSA has the potential to ignore (one-way) or exaggerate (extreme scenarios) the interaction between parameters, and provide results which are easy to misinterpret. Especially for the case of a multi-way sensitivity analysis, the mix of parameters to vary in combination and their possible relation can become complicated (Gray, Clarke

et al. 2010). As the choice of variables is dependent on arbitrary decisions (e.g. on choosing which variables to vary and in what range) DSA can become heavily sensitive to the discretion of the analyst. There is often a possibility that DSA becomes a tool providing estimates unrepresentative of the true uncertainty (i.e. over- or underestimate uncertainty), rather than providing a useful indication on the likelihood of model results.

Some limitations of DSAs, especially their limited ability to show joint parameter uncertainty and interactions between parameters, can theoretically be overcome by conducting a probabilistic sensitivity analysis. If the distribution around and the correlation between parameters is correctly specified, the PSA will provide a more precise estimation of mean costs and effects (Groot Koerkamp, Weinstein et al. 2010). Concerns expressed about PSA relate mostly to practice. Assumptions on the inter-dependence of parameters are rarely made and the choice of parameter distribution can sometimes be inappropriate. Once the analyst has to choose distributions and related parameters in an arbitrary fashion (e.g. due to the lack of data) many limitations of DSA will still hold true for PSAs, too. Hence PSAs are most helpful when the distribution and correlation of parameters are well-specified.

## 5.3 Value of information analysis

Decision-makers, with the authority to delay decisions or revisit them later, are interested not only in the expected outcomes and the uncertainty around the results, but also in the value of carrying out additional research. The results of the PSA help determine the value of acquiring additional information for future research by conducting the so-called value of information (VOI) analysis.

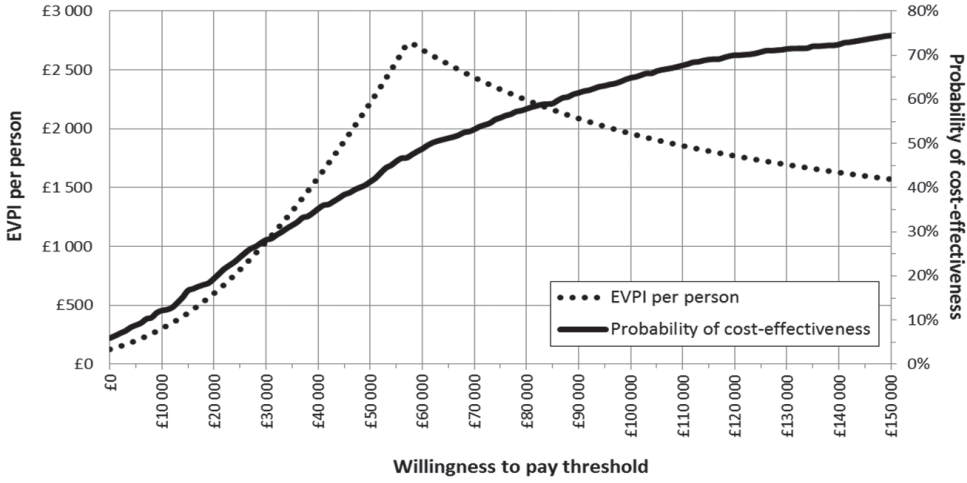
The idea of VOI is that information that emerges after the point of decision may have influence on changing the conclusion of the analysis (Griffin, Claxton et al. 2011). Such additional information can be drawn from future quantitative research – e.g. randomized controlled trials, observational studies. Due to additional information from new studies, parameter uncertainty is reduced and reimbursement of suboptimal interventions may be avoided, thus social welfare increases (Briggs, Claxton et al. 2006).

A VOI analysis relies on the assumption that at the time of the analysis we make the decision based on our current estimate of expected benefit. If our decision based on current information turns out to be wrong there will be costs in terms of health benefit and resources forgone. The expected cost of uncertainty is determined jointly by the probability that a decision based on existing information will be wrong and the consequences of a wrong decision (Briggs, Claxton et al. 2006). With estimates on the probability of error and the opportunity cost of error we can calculate the expected cost of uncertainty or the expected opportunity loss surrounding the decisions. The expected cost of uncertainty can be interpreted as the expected value of perfect information (EVPI), as perfect information

would eliminate the possibility of making the wrong decision. So, EVPI is the expected opportunity loss that could be avoided with perfect rather than current information. It can be simply estimated at any given willingness-to-pay threshold value for a unit of effectiveness, given the probability distribution of net benefit for the intervention (Eckermann, Karnon et al. 2010). In other words, EVPI is the difference between the payoff (costs vs. benefits) with perfect and current information. Further discussion and numerical illustration of the problem is provided at Briggs, Claxton et al. (2006).

The example on Figure 15 shows the relationship between the willingness-to-pay thresholds and the expected value of perfect information. If the EVPI (per person) exceeds the expected costs of additional research, then it is potentially cost-effective to conduct further research. For example, if additional research is expected to cost 1,500 pounds, then additional research will be potentially cost-effective when the threshold is greater than 40,000 pounds per QALY (Figure 15). So, it makes sense to reduce uncertainty for the price of 1,500 pounds/person[13]. At lower values of the threshold, e.g. 30,000 pounds per QALY, further research to reduce uncertainty should only be conducted if it does not cost more than 1,000 pounds (EVPI < 1,000 pounds/person).

**FIGURE 15. RELATIONSHIP BETWEEN A COST-EFFECTIVENESS ACCEPTABILITY CURVE AND THE EVPI FOR ONE PERSON**




As Briggs and Claxton et al. (2006) argue, the relationship between the EVPI and the cost-effectiveness threshold (shown on in Figure 15) has an intuitive interpretation. When

13 It is important that EVPI is also expressed for the total population of patients who stand to benefit from additional information over the expected lifetime of the technology. Hence to get the precise estimate of the total value of expected research per person, EVPI should be multiplied by the number of people affected by the treatment. This is called the population EVPI.

the threshold of CE is low, the technology is not expected to be cost-effective and additional information is unlikely to change that decision (EVPI is low). Thus any current evidence can be regarded as sufficient to support the decision to reject the technology. In this segment of the analysis (low threshold) the EVPI increases with the threshold because the decision uncertainty (probability of error) increases and the consequences of decision error (opportunity loss) are valued more highly. Conversely, when the threshold is higher than the incremental CE ratio (probability of cost-effectiveness is beyond 50%), the intervention is expected to be CE and this decision is less likely to be changed by further research as the threshold is increased. The decision uncertainty falls again because the technology appears to be increasingly cost-effective as the threshold is increased (probability of error falls, tending to reduce the EVPI).

In this example the reduction in decision uncertainty (threshold beyond app. 58,000) offsets the increased value of the consequences of error (EVPI). The EVPI reaches a maximum when the threshold is equal to the expected incremental cost-effectiveness ratio of this technology. In other words, the EVPI for this case reaches a maximum when we are most uncertain about whether to adopt or reject the technology based on existing evidence.

The wider application of EVPI analysis is currently under development (e.g. for resource allocation purposes) and is expected to become more widespread along with the standardized use of probabilistic sensitivity analyses (see in section 5.2.2) for economic evaluations. For a discussion on the relationship across certain derived measures of uncertainty, including VOI measures, see Campbell et al. (2015).

# 6 Validation of decision models

Balázs Nagy, László Szilberhorn and Anett Molnár

Validation, as discussed in section 4.2, is a key and integral part of the model development process. It demonstrates and evaluates whether the model is a proper and sufficient representation of the system under assessment and whether the results of the analysis can serve as a solid basis for decision-making (Vemer, Ramos et al. 2016). The results of the validation determine whether the model can be used by decision-makers to draw conclusions. On the other hand, validation is not about transparency. Transparency can help readers understand what a model does and how it does it, while validation determines how well the model serves the particular decision (Eddy, Hollingworth et al. 2012). The literature on validation in healthcare decision-modeling is rather immature, and the development of methodologies for validation (including quality assurance) is not fully elaborated. So far the focus has been on both the hard numbers and the softer processes of model development and problem structuring (Brennan and Akehurst 2000, Vemer, Ramos et al. 2016); theoretical and practical rules of thumb have equally been explored.

## 6.1 Common methodological flaws

Irrespective of the preparedness of the developer healthcare decision models may suffer from various types of flaws. Some of these might be related to computational errors, others might be due to inappropriate use of data, misinterpretation of real-world phenomena, or inappropriate choices about the methodology. Models always have potential to (adopted from Drummond and Schulpher [2005]):

- inadequately translate clinical data to economically justifiable data,
- inappropriately extrapolate beyond the observed period,
- choose wrong comparators,
- omit important costs or benefits,
- use assumptions excessively rather than using data,
- inadequately characterize uncertainty,

– show problems in the aggregation of results,
– selectively report on findings.

It is important for analysts and users to know how to detect these flaws. However, seeking the perfect model is not necessarily optimal. Most, if not all, reimbursement decisions are made at a time when full information is not available. Therefore the appropriate way to judge economic evaluations is not whether they embody some ultimate "truth," but whether they lead to a better decision than would have been made in their absence (Drummond and Sculpher 2005).

## 6.2 Types of validity

The extent to which a model is good – i.e. represents the reality it intends to describe – can be discussed by using the validity terms as follows (based on Eddy, Hollingworth et al. 2012).

**Face validity** is the extent to which a model, its assumptions, and applications correspond to current science and evidence, as judged by people who have expertise in the area. Four aspects are particularly in focus with face validity:

– model structure,
– data sources,
– problem formulation, and
– results.

In practice face validity is tested by people who have clinical or other health-related expertise to evaluate how well each component reflects their understanding of the pertinent medical science, available evidence, and practice. Consequently face validity is subjective in the sense that it reflects a qualitative analysis of selected experts and their judgement based on their understanding on the model and reality. It is still regarded as a crucial part of model building and inevitable for the acceptance of the final model.

**Internal validity** (also referred to as internal consistency, and technical validity) examines the extent to which the mathematical calculations are performed correctly and are consistent with the model's specifications. There are two main steps to check internal validity: i) verifying the individual equations and ii) checking their accurate translation to codes or formulae. Equations and parameters should be validated against their sources. Coding accuracy should be checked by using state-of-the-art quality assurance (see in section 6.4) and control methods for software engineering. These methods strongly depend on model complexity.

**Cross-validity** is checked when a model is compared with other models to determine the extent to which the models calculate similar results. These methods[14] examine different models that address the same problem and compare their results where any differences among results and underlying causes are examined.

**External validity** is checked when a model is used to simulate a real scenario, such as a clinical trial or an observational study, and the predicted model outcomes are compared with the real-world outcomes. Model results are compared with actual event data, frequency of clinical outcomes or other measures. It also involves simulating events that have occurred, such as those in a clinical trial, and examining how well the results correspond.[15] External validation applies to the model as a whole or to some components, such as population creation, disease incidence, disease progression, care processes and behaviors, occurrence of clinical outcomes, as well as interventions and their effects.

**Predictive validity** involves using a model to forecast events and, after some time, comparing the forecasted outcomes with the actual ones. Validation often involves recognizing a study design, simulating that design, recording the predicted outcomes, waiting for events to unfold, and comparing them with predictions. This process is frequent in the case of clinical trials that have published their designs, but not yet reported results; it can also be applied to cohort studies still in progress.

## 6.3 Phases of model validation

While validation can be identified in the sequence of model development (see section 4.2) it has a distinct relation to all phases, too (see Figure 16). Different types of validation techniques can be linked to each phase of the model development (Sargent 2005):

1. **Conceptual model validation** determines that the theories and assumptions underlying the conceptual model are correct and that the model's structure, logic, mathematical and causal relationships are reasonable.
2. **Computerized model verification** assures that the computer programming and implementation of the model concept is correct. It entails techniques which ensure that the implemented software program including code, mathematical calculations and implementation of the model concept are performed correctly and are consistent with the model's specifications.
3. **Data validation** encompasses techniques used to determine whether the available input data is appropriate, accurate and sufficient and that data transformations

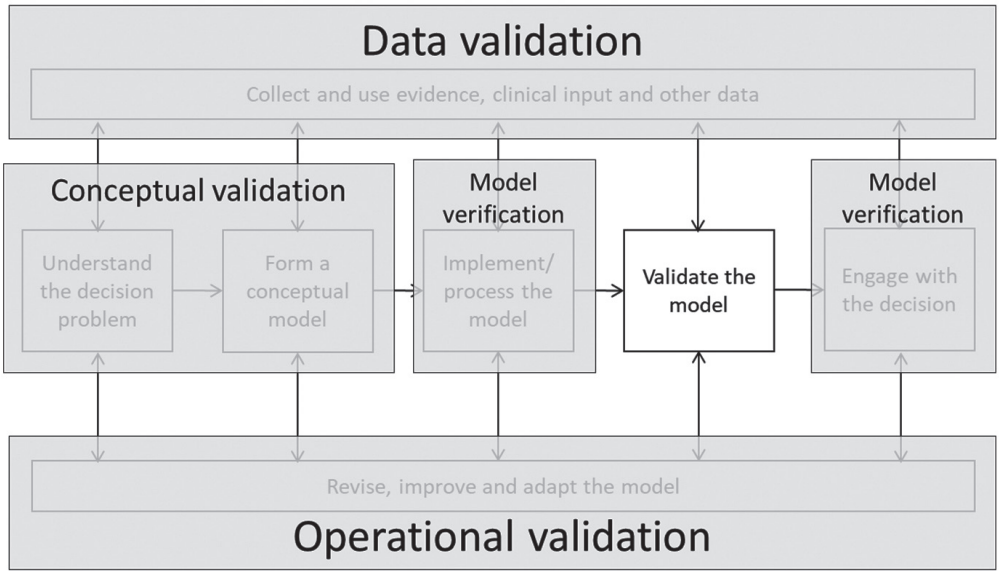
14 These methods are also called external consistency, comparative modeling, external convergence testing, convergent validity, external consistency and model corroboration.

15 For multi-application models external validations can be applied to the model in a general sense and to each application. It is important to perform multiple validations that crisscross the intended applications in the sense of involving a range of populations, interventions, outcomes, and time horizons.

were correctly performed. This process ensures that the data necessary for model building, model evaluation and testing and conducting the model experiments to solve the problem are adequate and correct.

4. **Operational validation** determines that the model outputs have sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability. It practically validates the real life case against a specific model application and implementation.

**FIGURE 16 VALIDATION TECHNIQUES AND THE PHASES OF MODEL DEVELOPMENT**




*Note:* phases of model development were originally detailed on Figure 12

Although not presented in Figure 16, the validation tasks strongly overlap and are often repeated until the model itself reaches a satisfactory level of quality and acceptability. Usually dozens of model versions are developed prior to obtaining the valid model.

Validation types are usually conducted in the context of specific applications. A model can have different levels of validity for different applications and the validation may apply to particular applications, not to the model itself.

The required degree of validity depends on the research context: e.g. predicting 30-year versus 6-month survival probabilities for oncology therapies requires very different levels of accuracy. Consequently, it is difficult to specify criteria a model must meet to be declared "valid," as if validity were a property of the model that applies to all its applications (Eddy, Hollingworth et al. 2012).

No matter how many validations are done, there will inevitably be uncertainty about some aspects of a model. A sensitivity analysis can be used to explore how a model's results change with variation to inputs; but on its own, it will not evaluate how accurately a model simulates what occurs in reality so it cannot substitute for validation.

Table 5 in Appendix VI. gives a comprehensive overview of the relationship between phases of model development, validation techniques and the type of validity which the modeler should check.

## 6.4 Tools of model validation

Plenty of validation tools are available in the public domain to help model developers and users to test and confirm their findings. The latest review by Vermer and colleagues (2016) reported 31 checklists in practice, all of which are potentially helpful for different purposes. The applicability of validation tools depends on the context, time and resources available, and also on the significance of the decision to be made.

Currently there is no predefined set of tools or methods which will confirm that a health economic decision model is valid. Choices for the modeler and the user are strongly driven by a series of individual decisions throughout the model building and implementation process.

Validation along the model development process is presented by the checklist developed by the authors of this book. This guide sets up links between model building activities and phases of validation for various types of modelling exercises. As seen in Appendix VI. the process and tasks become fairly technical and detailed once it comes to real-life practice. Also types of models, their level of difficulty and time constraints may lead to different sets of activities.

An example on technical validation is also provided in Appendix V. This checklist presents tasks which are useful for modelers and skilled users to test if the model calculations are correct (model verification). Again it is the discretion and responsibility of the modeler/user to carry out the suggested types of activities to an extent which is deemed necessary. It is also important to document the validation process, which endorses the transparency and the replicability of the model.

# Appendix I

Based on the study of Brennan et al. (2006) this section describes Table 1 (presented in section 2.2) in detail.

Deterministic decision trees (**A1**) are widely used aggregate level models for healthcare decision analyses. They outline the structure of decision problems, the probability or fraction of various outcomes of the decisions and the valuation of their outcomes (by using e.g. quality-of-life, cost or net benefit measures). The mean value of a decision is computed by summing the probability of each outcome multiplied by its value (see more in section 3.1). Stochastic cohort models (**column B**) assume randomness. The simulated decision tree cohort (**B1**) provides an alternative approach to the deterministic decision tree: it simulates the number of individuals on each path of a decision tree (but not each individual) separately to get an idea of the variability around the mean results. Markov models for cohorts can be processed either analytically (**A2**) with expected values or with simulated random Markov model transitions (**B2**) – the same way as simulated decision tree cohorts. The advantage of the stochastic approach is, similarly to **B1**, that it can provide a measure of the variability of the number of individuals likely to be in each state of the cohort.

Individual level models (ISM) (**columns C, D**) simulate the progression of each individual with different characteristics. Rather than tracking data for every pathway (as cohort models do), ISMs track specific individuals and generate large numbers of simulated patient histories to evaluate results (see more in sections 3.4 and 3.5). Simulated Patient Level Decision Tree models (**CD1**) take individuals through a tree's different pathways and make a record of the patients' (disease/treatment) history. Similarly, individual Markov simulation models take individuals with certain characteristics through model states in each time period (**CD2**). By tracking multiple co-morbidities or other attributes, they are able to depict complex diseases/situations. Examples include models of diabetes where patient co-morbidities interact and affect the outcomes, and other models on rheumatoid arthritis and osteoporosis (see more in section 3.4).

Amongst models allowing interactions between individuals, system dynamic models (**A3, A4**) are of great importance in public health, epidemiology and operational research. However, they are less frequently used for economic evaluation in healthcare. With the system dynamics approach the state of the system is modelled in terms of changes over time. The underlying philosophy is that if individuals do interact, by using this approach, a cohort based model can still be sufficient; it is not necessary to model each individual separately. These models assume that the rate of change in the system is a function of the system's state itself (i.e. feedback). Feedback loops are closed circles of causal relationships that work to amplify or resist changes introduced into systems. They have a critical impact on the behavior of complex systems. Typical examples of feedback include infectious disease outcomes, where higher levels of infection produce higher risks of fur-

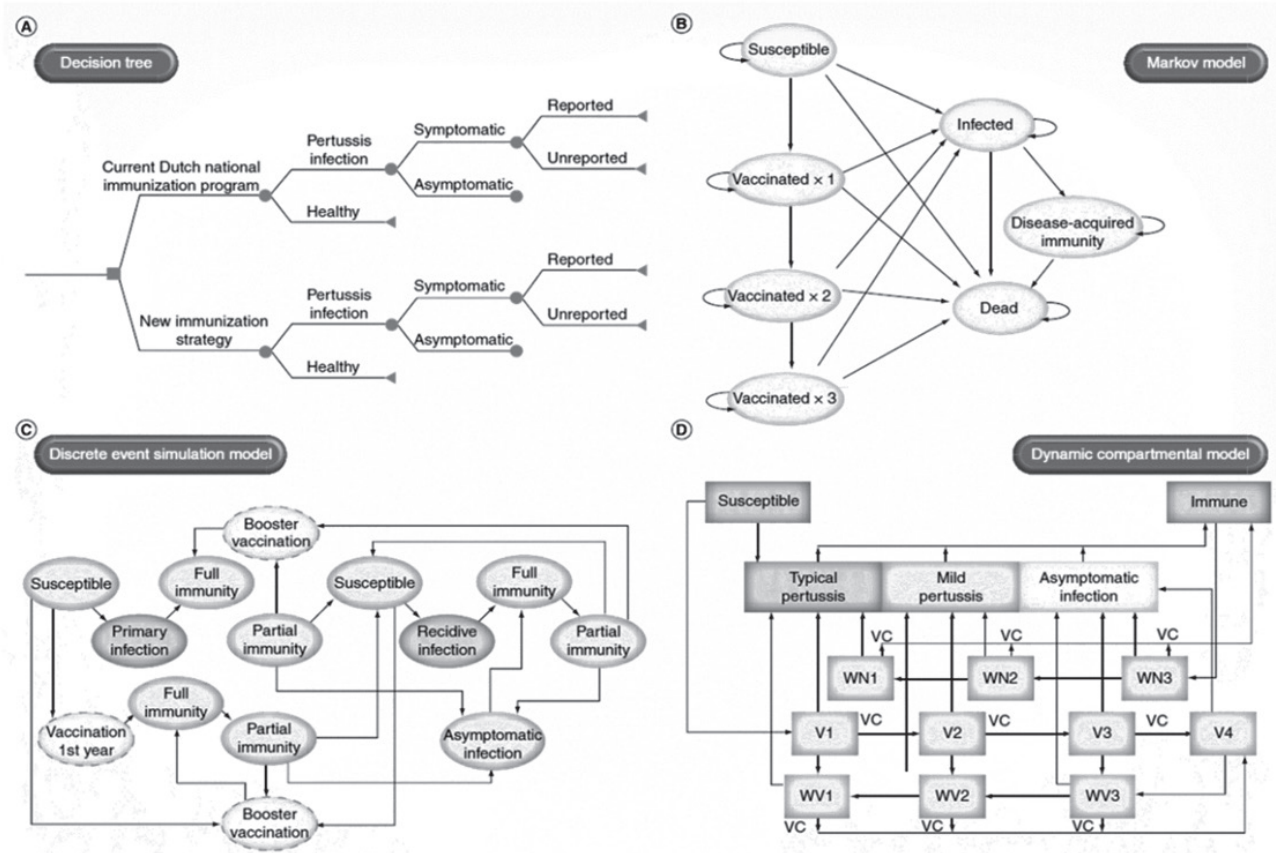
ther infection, or healthcare service constraints, where the system performs differently when it is full or over-capacity. Typical applications of dynamic models are concerned with community care, short-term psychiatric care, HIV/AIDS (Dangerfield 1999), smallpox preparedness, Creutzfeldt–Jakob disease (Bennett, Hare et al. 2005).

The system dynamics models are based on two important assumptions: i) the change dynamics are *set deterministic* and ii) fractions of individuals can occur in specific health states (i.e. an infinitely divisible population assumption). However, these assumptions are only feasible when large populations are analyzed. To address the variability of the system dynamics and the need to model integer numbers of individuals in each health state, continuous time Markov chain (CTMC, **B4**) models can be applied. A CTMC can jointly model many interactions including, for example, infectious disease dynamics and limited healthcare resources. These models are often analyzed stochastically to describe how the state changes through time. The number of individuals in each state is tracked, and each individual event must be processed, but there is no need to sample the identities of individual patients to maintain their histories. The discrete time Markov chain (DTMC) model (**B3**) is like the CTMC except that it is updated with finite time steps.

Individual level models with interactions allow even more flexibility and heterogeneity. Individual level Markov models (**C3**, **C4**) with interactions can be thought of as an extension of Markov cohort models with interaction (**B3**, **B4**). Individual level means that patient characteristics may be heterogeneous, and that histories may be tracked for each individual in the population. These are also called individual event history (IEH) models. In continuous time IEH models (**C4**) the parameters and rates may differ for each individual to reflect heterogeneous population characteristics, and rates may also depend upon resource constraints (as a result of taking into account interactions between individuals). The analogous discrete time IEH model (**C3**) steps *forward in* discrete time intervals in the same way as the DTMC model (see previous paragraph).

Models with non-Markovian properties (**column D**) allow greater flexibility in modelling the timing of health-related events (i.e. the time is not constrained by fixed model cycles). Complex individual level models (e.g. **D4**) can examine interactions both with other individuals and with the environment, including the availability of resources (e.g. doctors, beds, surgery room, transplant organs). Probably the most flexible of all modelling techniques is continuous time discrete event simulation (CT, DES) (**D4**). It describes the progress of individuals (entities), which undergo various processes (events) that affect their characteristics and outcomes (attributes) over time. In these models the line structure (e.g. considering the event of other individuals when lining up for a surgical intervention) enables interaction to take place with constraints and between entities. For DES models the state of the modelled system includes the current entities, their attributes, and a list of events that can occur either at the current simulation time or that are scheduled to occur in the future (see more in section 3.5). Continuous time DES models have discrete-time analogues (**D3**) that are not different in any significant methodological way from the continuous time models if the discrete time steps are small enough.

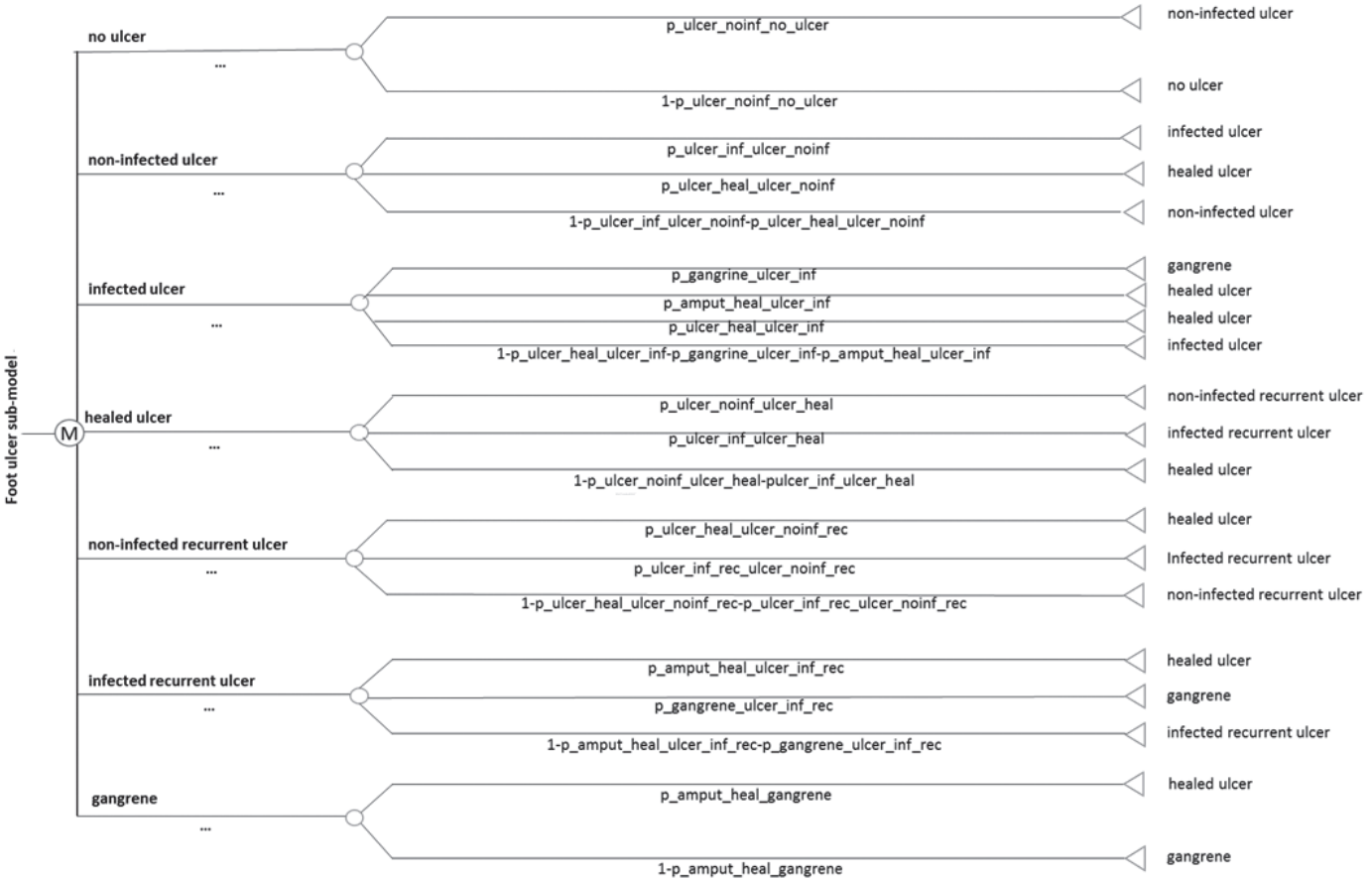
**FIGURE 17 FOUR MODELS TO ANALYZE THE COST-EFFECTIVENESS OF PERTUSSIS IMMUNIZATION STRATEGIES: A) DECISION TREE MODEL, B) MARKOV MODEL, C) DES MODEL, D) DYNAMIC MODEL.**




Key: V: Vaccine status; VC: Vaccination coverage; WN: Waning of natural immunity; Wv: Vaccine waning.

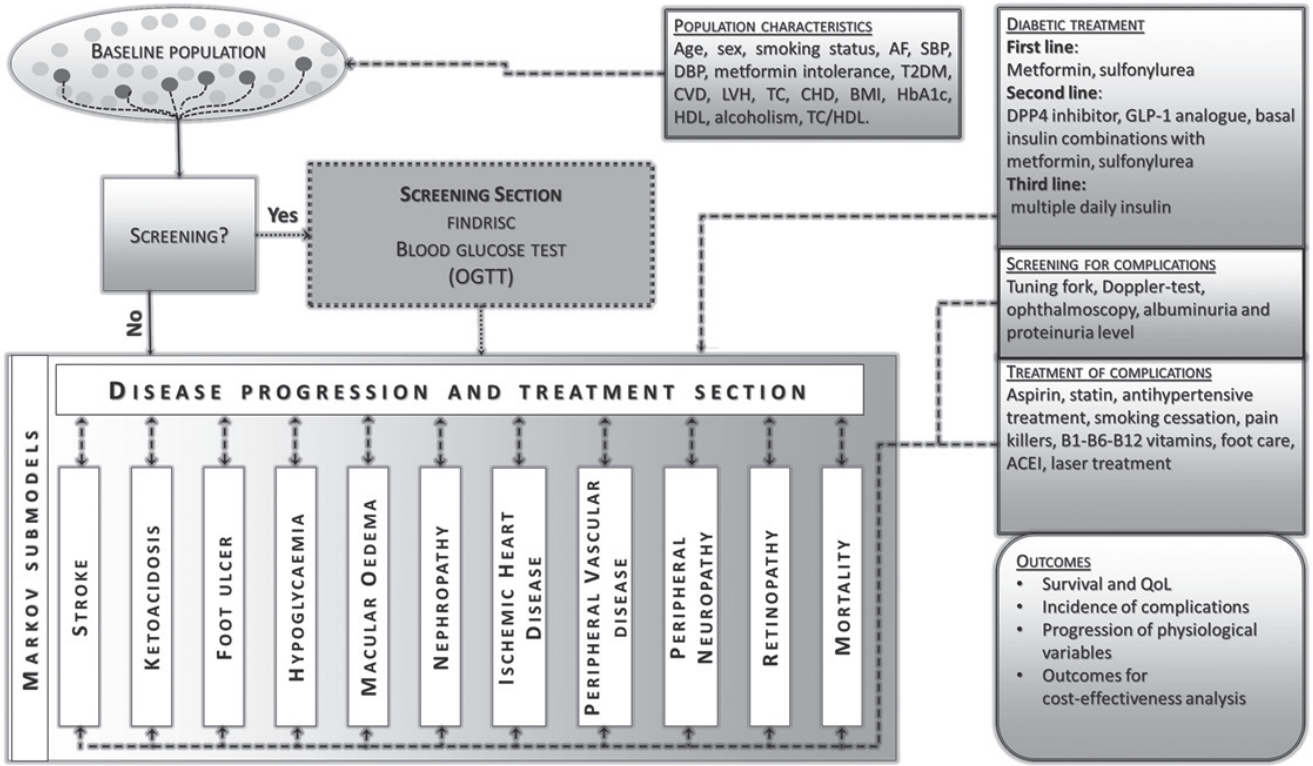
Source: Millier et al. (2012)

**FIGURE 18 MARKOV TREE TO DESCRIBE THE STRUCTURE OF A MARKOV MODEL**



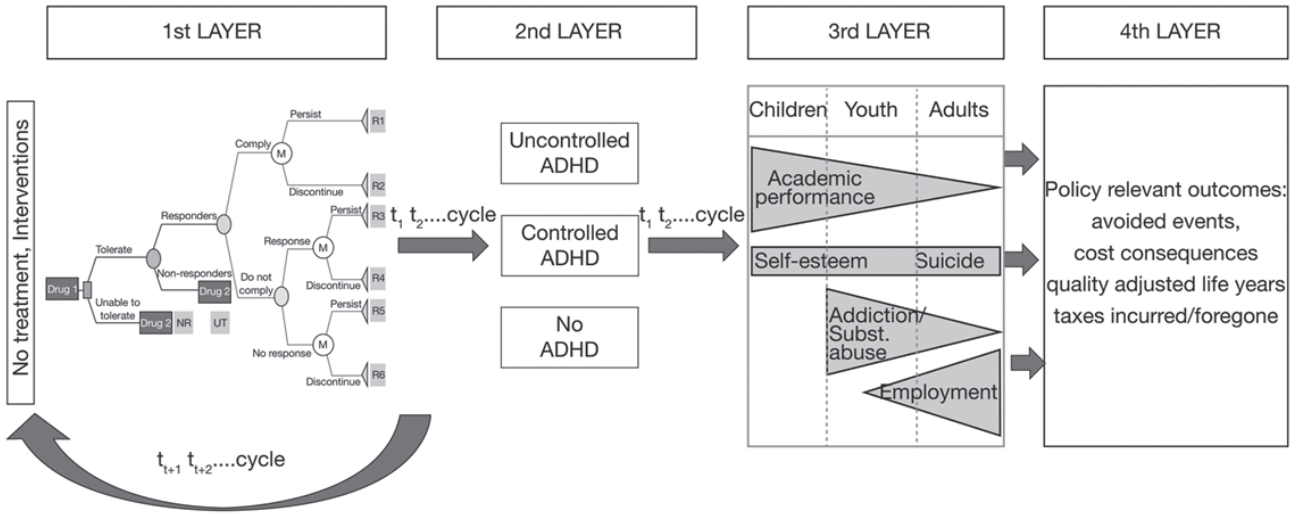

Source: based on Nagy et al. (2016)

**FIGURE 19 STRUCTURE OF THE SYREON TYPE 2 DIABETES MODEL**




Source: Nagy et al. (2016)

**FIGURE 20 STRUCTURE OF THE LONG-TERM ECONOMIC MODEL OF ATTENTION-DEFICIT HYPERACTIVITY DISORDER**




Source: Nagy et al. (2017)

# Appendix V

## Technical validation checklist for decision analytic models

Source: Syreon Research Institute, Modeling Division, 2017

This series of analyses help check the technical correctness of decision analytic models (mainly built in Excel). Not all tests are applicable for each model, hence the use of all tests is not expected.

**TABLE 5**

| | | ICER | Expected effect | Effects on the model, explanation | Decision (OK / NO) |
|---|---|---|---|---|---|
| Default value | | | | | |
| General input settings | Set costs discount rate to 0% | | Same costs as undiscounted rate | | Ok |
| | Set benefits discount rate to 0% | | Same benefits as undiscounted rate | | Ok |
| | Check both arms' sums to cohort size plus deaths each year | | Should sum to cohort size* if applicable | | Ok |
| | Check the numbers of patients in each of the health states are the same for each treatment | | | | NA |
| | Set all efficacies / hazard ratios/ withdrawal rates/ mortality/ RR etc. to the same figure (0 and/or 1) | 0 | Same number of QALYS for both arms AND/OR all incremental QALYs = 0 / 1 | Setting market share of INF to almost 100%, CDX +ve response to 100% and Multiplier for increased efficacy to 1 gave the same QALYs and costs | Ok |
| | | | | | |
| This test demonstrates that all patients alive are being counted in the LYG and QALY calculations & that QALYs calculations are correct. | Set discount rate for utilities to zero. Set all utilities for states to 0 and utility decrement vales to 0 and set mortality to 0. | | Total QALY = 0 | Set AE disutility =0<br>Set utility intercept and all factors = 0<br>Set mortality = 0<br>Discount rate= 0 (result 0 QALYs) | Ok |
| | Without changing the previous settings change utilities for states to 1 | | Total QALY = model time horizon | Then changed intercept to 1 (result 10 QALYs) | Ok |
| | | | | | |

| | | ICER | Expected effect | Effects on the model, explanation | Decision (OK / NO) |
|---|---|---|---|---|---|
| This test demonstrates that differences in patient's numbers in each of the health states between treatments are due to differences in progression free and/ or overall survival. | Set each of the cost parameters to 0 | | Total cost is zero | | Ok |
| | Check that the total costs for both treatment arms are 0 for every year | | Costs both arms =0<br>QALYs give values for both arms | | Ok |
| | Check that the total cost (in the results table) is 0 for both treatment arms | | | | Ok |
| | | | | | |
| Transitional probabilities | Check transition matrices e.g. setting same values in the matrices for both arms | | | The comparator arm is not similar in structure to the new treatment arm | No |
| | Check sum of rows in each transition matrix | | All rows should sum to 1 | | NA |
| | | | | | |
| This test ensures costs calculations are correct | Are all health states accounted for regarding costs? | | | | Ok |
| | Are costs adjusted to cycle length? | | | | Ok |
| Benefits | Are all health states accounted for? | | | | Ok |
| | Are benefits adjusted to cycle length | | | | Ok |
| This test demonstrates that the two treatment arms have the same logic and calculations, and hence will come to the same results with the same data | Ensure that all values in every cell of 'Data entry' and 'Parameter' sheets are identical, using a comparison sheet. | | ICER = #DIV/0<br>Total cost =<br>Total QALY = | | NA |
| | Ensure differences in totals in 'Data entry' and 'Parameter' sheets are zero | | In case where data entry sheet is not used directly for calculations | | Ok |
| | Ensure all in differences in CE calculations are zero, and CE ratios are n/a | | | | |
| | Set all PFS/OS γ and λ values to equal for both arms | | Not applicable in all models | | NA |
| This test ensures that patients are neither entering nor leaving the model – only changing from one state to another. | | | | | |
| | Check sum of all health states =100% at all times | | | | Ok |

| | | ICER | Expected effect | Effects on the model, explanation | Decision (OK / NO) |
|---|---|---|---|---|---|
| Further testing can be conducted using a range of extreme parameter values, and the structure, cell formulae and visual basic code was examined for programming errors. | Check that the sum of each year's patient flow (in both arms) sum to 100% | | | | Ok |
| Checking terminal states | Check that the number of patients in terminal states is not less than it was in the previous cycles | | | | Ok |
| | | | | | |
| Checking Discounting Formulas for Costs and Benefits | =1/(1+Discount rate)^(t-1) | [redacted] | | | Ok |
| | Effective rate for period = (1 + annual rate)^(1 / # of periods) – 1 | | | | NA |
| | | | | | |
| Inputs | Are values from input sheet used in calculations? | | It may be direct or indirect, but there must be a link between each single cell in input sheet and model engine – check the VBA code if necessary | | Ok |
| Results | Costs are referenced to the correct cell | | Reference cells are consistent with the model length, and are referencing to the same cycle, or same logic – check the VBA code if necessary | | Ok |
| | QALYs are referenced to the correct cell | | | | Ok |
| Adherence | Applied to Costs? | | It should be the same for Costs and Benefits | | Ok |
| | Applied to Benefits? | | | | Ok |
| VBA | Check if the model uses VBA code and examine the calculations which are done in VBA. | | | | NA |

| | | ICER | Expected effect | Effects on the model, explanation | Decision (OK / NO) |
|---|---|---|---|---|---|
| PSA | Apply Syreon PSA Template with same variables included in the original model and same distributions chosen | | Results should be nearly identical | Originally developed using Syreon PSA template. | NA |
| | Set all inputs that affect utilities not to be included in PSA. | | PSA plot should be a line parallel to the cost axis (Y axis most probably) | | NA |
| | After the previous step, start switching on the variable groups one by one. | | PSA plot should look healthy otherwise the recently switched on variable group is causing a problem. | | NA |
| | Set all inputs that affect costs not to be included in PSA | | PSA plot should be a line parallel to the QALYs axis (X axis most probably) | | NA |
| | After the previous step, start switching on the variable groups one by one. | | PSA plot should look healthy otherwise the recently switched on variable group is causing a problem. | | NA |
| DSA | Set upper multiplier to be 5 times the lower multiplier | | Results for decreasing value should be negligible compared to increasing it | | Ok |
| | Set lower multiplier to be 5 times the upper multiplier | | Opposite | | Ok |

## TABLE 6 INTERNAL QUALITY ASSURANCE CHECKLIST FOR MODELERS TO TEST THE MODEL ALONG THE PHASES OF DEVELOPMENT

| Development phase to validate | validation task | action of validation | outcome of validation | | | | | | contributor | | | | | Estimated time in weeks | | | | Phase of validation | | | | Validity to check | | | | | Type of model to check | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | comment with date | technical checklist | paragraph in report | new model | new report | | model developer | internal QA modeller* | external QA modeller** | modelling supervisor | clinician | 0.5-1.5 | 1-2 | 3-5 | 4-6 | conceptual | computerized | data | operational | face | internal | cross | external | predictive | complex simulation model | simple simulation model | complex cohort model | simple model or adaptation | budget impact model | |
| **Model conceptualization** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| form model concept | examine the model concept in the light of available data | review the model structure and evidence | ■ | | | | | | ■ | | | ■ | ■ | ■ | | | | ■ | | | | ■ | | | | | ■ | ■ | ■ | ■ | ■ | |
| **Technical model development** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| build modelling tool including formulas, codes | compare programed model structure with model concept | review excel spreadsheet | ■ | | | | | | ■ | | | ■ | | ■ | | | | | ■ | | | | ■ | | | ■ | ■ | ■ | | | ■ | |
| incorporate value drivers | check and validate the critical inputs/assumptions | critical input/assumption review and supporting analysis when needed | ■ | | | | | | ■ | | | ■ | ■ | ■ | | | | | ■ | ■ | | ■ | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | |
| populate the model with parameters | compare parameters with sources | review model parameters and sources | ■ | | | | | | | ■ | | | | ■ | | | | | | ■ | | | ■ | | | ■ | ■ | ■ | ■ | ■ | |
| draft model | check formulas and results | run technical checklist preview results | ■ | ■ | | | | | | ■ | | | ■ | | ■ | | | | ■ | | | ■ | ■ | | | ■ | ■ | ■ | | | |
| user interface | check functionality of formulas | review formulas and functionality | ■ | | | | | | | ■ | | | | | ■ | | | | ■ | | | ■ | ■ | | | ■ | ■ | ■ | | | ■ |
| deterministic sensitivity analysis | check functionality and face validity of results | review formulas, programming and results | ■ | | | | | | | ■ | | | ■ | | ■ | | | | ■ | | ■ | ■ | ■ | | | ■ | ■ | ■ | | | ■ |
| probabilistic sensitivity analysis | check functionality and face validity of results | review formulas, programming and results | ■ | | | | | | | ■ | | | | | ■ | | | | ■ | | ■ | ■ | ■ | | | ■ | ■ | ■ | | | |
| **Reporting** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| draft report +final model 1 | write up the report and check all details in the model | check as you write up the report | ■ | | | | | | ■ | | | ■ | | | ■ | | | | | | ■ | ■ | ■ | | | ■ | ■ | ■ | ■ | ■ | ■ |
| draft report +final model 2 | compare report with the model | check document with model +run technical checklist | ■ | ■ | | | | | | ■ | | | | | ■ | | | | | | ■ | ■ | ■ | | | ■ | ■ | ■ | | | ■ |
| final report with final model interface | review report and model from a user perspective | check usability and readiness of model and document | ■ | | ■ | | | | | | ■ | ■ | | ■ | | | | | | | ■ | ■ | | ■ | ■ | ■ | ■ | | ■ | ■ | |
| **Optional phases** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| rebuilt model with alternative technique | rebuild the model using an alternative modelling technique | build new model | | | | | ■ | ■ | ■ | | | ■ | | | | | ■ | | ■ | | ■ | | | ■ | | ■ | ■ | | | | |
| reprogrammed model with alternative software | reprogram the model using another software | build same model with other program | | | | | ■ | ■ | | ■ | ■ | | | | | ■ | | | ■ | | | | ■ | | | ■ | | ■ | | | |
| final report with externally valid model | check validity of results based on empirical evidence and external clinician input | use the existing model to make comparisons with empirical evidence | | | | ■ | | ■ | ■ | | | ■ | ■ | | ■ | | | | | ■ | ■ | | | | ■ | | ■ | ■ | | | |

* can be included in model development
**not involved in model development
Comment: approval or rejection with minor or major corrections or any other issues
During adaptation we assume that the model structure and functionality is good

# References


Ágh, T., et al. (2016). "The cost effectiveness of lisdexamfetamine dimesylate for the treatment of binge eating disorder in the USA." *Clin Drug Investing* **36**(4): 305–312.

Andronis, L., et al. (2009). *Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making*, Prepress Projects Limited.

Araújo, D. V., et al. (2008). "Cost-effectiveness of prehospital versus inhospital thrombolysis in acute myocardial infarction." *Arquivos Brasileiros de Cardiologia* **90**(2): 100–107.

Basharin, G. P., et al. (2004). "The life and work of AA Markov." *Linear Algebra and its Applications* **386**: 3–26.

Bennett, P., et al. (2005). "Assessing the risk of vCJD transmission via surgery: models for uncertainty and complexity." *Journal of the operational research society* **56**(2): 202–213.

Berger, M. L., et al. (2003). *Health Care Cost, Quality, and Outcomes. ISPOR Book of Terms*, Lawrenceville NJ.

Brennan, A. and R. Akehurst (2000). "Modelling in health economic evaluation." *Pharmacoeconomics* **17**(5): 445–459.

Brennan, A., et al. (2006). "A taxonomy of model structures for economic evaluation of health technologies." *Health economics* **15**(12): 1295–1310.

Briggs, A., et al. (1994). "Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis." *Health economics* **3**(2): 95–104.

Briggs, A. H., et al. (2006). *Decision modelling for health economic evaluation*, Handbooks in Health Economic E.

Briggs, A. H., et al. (2012). "Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6." *Value in health* **15**(6): 835–842.

Brisson, M. and W. Edmunds (2003). "Economic evaluation of vaccination programs: the impact of herd-immunity." *Medical Decision Making* **23**(1): 76–82.

Bródy, F. and T. Vámos (1995). *The Neumann compendium*, World Scientific.

Campbell, J. D., et al. (2015). "Cost-effectiveness uncertainty analysis methods: a comparison of one-way sensitivity, analysis of covariance, and expected value of partial perfect information." *Medical Decision Making* **35**(5): 596–607.

Caro, J. J., et al. (2010). "Discrete event simulation: the preferred technique for health economic evaluations?" *Value in health* **13**(8): 1056–1060.

Chilcott, J., et al. (2010). "Avoiding and identifying errors in health technology assessment models: qualitative study and methodological." *Health Technology Assessment* **14**(25).

Consortium, Y. H. E. (2016) Budget Impact Analysis.

Dangerfield, B. (1999). "System dynamics applications to European health care issues." *Journal of the operational research society* **50**(4): 345–353.

Drummond, M. and M. Sculpher (2005). "Common methodological flaws in economic evaluations." *Medical care* **43**(7): II–5–II–14.

Drummond, M. F. and A. McGuire (2001). *Economic evaluation in health care: merging theory with practice*, OUP Oxford.

Eckermann, S., et al. (2010). "The value of value of information." *Pharmacoeconomics* **28**(9): 699–709.

Eddy, D. M., et al. (2012). "Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7." *Medical Decision Making* **32**(5): 733–743.

Gray, A. M., et al. (2010). *Applied methods of cost-effectiveness analysis in healthcare*, OUP Oxford.

Griffin, S. C., et al. (2011). "Dangerous omissions: the consequences of ignoring decision uncertainty." *Health economics* **20**(2): 212–224.

Groot Koerkamp, B., et al. (2010). "Uncertainty and patient heterogeneity in medical decision models." *Medical Decision Making* **30**(2): 194–205.

Group, S. F. U. C. S. M. (2010). *Modelling in Healthcare*, American Mathematical Society.

Heeg, B. M., et al. (2008). "Modelling approaches." *Pharmacoeconomics* **26**(8): 633–648.

International, H. T. A. (2017). What is HTA? Accessed Aug. 1, 2017. HTAi.

Karnon, J., et al. (2012). "Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–4." *Medical Decision Making* **32**(5): 701–711.

Mauskopf, J. A., et al. (2007). "Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices—budget impact analysis." *Value in health* **10**(5): 336–347.

Millier, A., et al. (2012). "A critical literature review of health economic evaluations in pertussis booster vaccination." *Expert Rev Pharmacoecon Outcomes Res* **12**(1): 71–94.

Nagy, B., et al. (2017). "A conceptual framework for a long-term economic model for the treatment of attention-deficit/hyperactivity disorder." *Expert Rev Pharmacoecon Outcomes Res* **17**(3): 283–292.

Nagy, B., et al. (2016). "Cost-effectiveness of a risk-based secondary screening programme of type 2 diabetes." *Diabetes Metab Res Rev* **32**(7): 710–729.

Németh, B., et al. (2017). "Quality-adjusted life year difference in patients with predominant negative symptoms of schizophrenia treated with cariprazine and risperidone." *Journal of comparative effectiveness research* **6**(8): 639–648.

Pitman, R., et al. (2012). "Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5." *Value in health* **15**(6): 828–834.

Roberts, M., et al. (2012). "Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force–2." *Medical Decision Making* **32**(5): 678–689.

Robinson, S. (2008). "Conceptual modelling for simulation Part I: definition and requirements." *Journal of the operational research society* **59**(3): 278–290.

Robinson, S. (2010). "Conceptual modelling: Who needs it." *SCS M&S Magazine* **2**: 1–7.

Sargent, R. G. (2005). *Verification and validation of simulation models*. Proceedings of the 37th conference on Winter simulation, winter simulation conference.

Sonnenberg, F. A. and J. R. Beck (1993). "Markov models in medical decision making: a practical guide." *Medical Decision Making* **13**(4): 322–338.

Tappenden, P. (2012). "Conceptual modelling for health economic model development." *HEDS Discussion Paper 12/05*.

Tappenden, P., et al. (2006). "Methodological issues in the economic analysis of cancer treatments." *European journal of cancer* **42**(17): 2867–2875.

Tesar, T., et al. (2017). "Cost-Utility Analysis of Heberprot-P as an Add-on Therapy to Good Wound Care for Patients in Slovakia with Advanced Diabetic Foot Ulcer." *Frontiers in pharmacology* **8**: 946.

Vemer, P., et al. (2016). "AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users." *Pharmacoeconomics* **34**(4): 349–361.

Vokó, Z., et al. (2017). "Az alacsony dózisú CT-vel végzett tüdőrákszűrés magyarországi egészség-gazdaságtani elemzésének koncepcionális terve." *Orvosi Hetilap* **158**(25): 963–975.

Weinstein, M. C. (2006). "Recent developments in decision-analytic modelling for economic evaluation." *Pharmacoeconomics* **24**(11): 1043–1053.

There are many high-quality health economic evaluation books and manuscripts. The objective of this book was not to be yet another health economic evaluation text. This book fills a void missing within alternative resources. The primary objective of its publication is to support eager learners and model building practitioners seeking a pragmatic and concise roadmap for how to choose wisely related to the many important decisions within health economic evaluation modelling.

This book begins with a practical review of decision analytic modelling techniques supporting economic evaluations in health care. After bringing learners and future and current model builders to an equal playing field, this book's essence relates to how it supports the reader in choosing a model type that fits the research question; in walking the reader through pragmatic step-by-step instructions for model building; in concisely addressing the advanced topic of uncertainty; and in providing checklists related to model validation and quality assurance.

Modelling, done well, is a rigorous, systematic, scientific exercise that transparently addresses research questions while simultaneously generating additional hypotheses.

The editors and authors combined scientific knowledge with many years of modelling experience to share the book's essence with readers who are interested in learning and working in this growing field.

ISBN 978-963-489-119-2