Ph.D. THESIS

2019

Cost-effectiveness analysis with new models of artificial intelligence. Medical applications

Jorge Pérez-Martín

Doctoral Program on Intelligent Systems

Academic advisor:

Francisco Javier Díez Vegas

Department of Artificial Intelligence
Universidad Nacional de Educacion a Distancia (UNED), Spain

2019

Assessment committee:

President Mª Elena Hernando Pérez

Secretary Cristóbal Belda Iniesta

Member Silvana Quaglini

Alternate of the president María Jesús Ledesma-Carbayo

Alternate of the secretary Antonio Sarria-Santamera

Alternate of the member Thomas D. Nielsen

ISBN XXX-XX-XXXXX-XX-X

To my fiancée and my family.
Thank you for your love, support and infinite patience.

Acknowledgments

Many people have helped and supported me during this personal and professional journey that has been my thesis. To those people, thank you for being here. Your presence has been my strength.

In the first place I wish to thank my PhD advisor, professor Francisco Javier Díez, for giving me advice and support not only for this thesis but for my personal and professional development. I am also grateful to the rest of the members of the CISIAD: Manuel Arias, with whom I have had fun and have discussed equally, and Manuel Luque, who has always been willing to dedicate a bit of his time to share his experiences. I want to thank the new members of CISIAD because with their camaraderie and good mood have enriched the work environment in the office.

On the other hand, I want to thank to the members of the miniXmodular group: Emilio Letón, whom I consider as a friend, his support and advice helped me to improve my teaching skills and his jokes are always appreciated; Alejandro Rodríguez Ascaso, with whom I have learned the importance of accessible design and I have been able to improve my digital materials; Elisa Mª Molanes, whose motivation has inspired me and whose knowledge has motivated me.

Finally, from a personal point of view, I wish to mention my family and my fiancée: my parents, Maria Jesús and Juan Bautista, who have supported me since I can remember; my brother Héctor, who has always been an example to follow and my sister-in-law Aurora, who supports him to remain so; my little niece Carmen, who brings joy and happiness to the family; and my fiancée Marta, whose love supports me everyday.

This thesis and the short stay at the University of Sheffield have been supported by a predoctoral grant (FPU14/02577) from the Spanish Ministry of Education. Our research has also been supported by the Health Institute Carlos III of the Spanish Government (grant PI13/02446), cofinanced by the European Regional Development Fund (ERDF), and by the Ministry of Science, Innovation and Universities (grant TIN2016-77206-R).

Abstract

Given the limited resources of health care systems, the economic evaluation of medical techniques and treatments is becoming more and more important. The most a common tools for cost-effectiveness analysis are decision trees and Markov models, but they can only solve relative small problems. Probabilistic graphical models, such as Bayesian networks and influence diagrams, have been used in artificial intelligence for knowledge representation and explanation, especially in medicine, but only in unicriteria problems. In the last years, the Research Center for Intelligent Decision-Support Systems (CISIAD) at UNED, has developed new algorithms for cost-effectiveness analysis with decision trees and influence diagrams. It has also proposed two new types of probabilistic graphical models: Markov influence diagrams which extend influence diagrams for temporal reasoning, and decision analysis networks, which can model and evaluate problems with asymmetries such as restrictions and partially ordered decisions.

This thesis addresses three methodological problems related to the evaluation of cost-effectiveness models.

First, there are different corrections for reducing the error introduced by the discretization of time in Markov models. In general, numerical integration techniques give more accurate results than standard approaches, such as half-cycle correction, but we found that they can lead to a greater error when the model has discontinuities, for example when an expensive treatment is withdrawn after some time. We proved that building a new model averaged at the points of discontinuity yields much more accurate results.

Second, the existing cost-effectiveness algorithms for Markov influence diagrams could only evaluate models with two criteria and one decision. In this thesis I have developed a new cost-effectiveness algorithm that can evaluate models with several criteria and any number of decisions with findings between the decisions.

Third, decision analysis networks could only evaluate unicriterion problems. In collaboration with other members of the CISIAD, I have extended the algorithms developed for Markov influence diagrams, to perform cost-effectiveness analysis on

decision analysis networks.

I have applied Markov influence diagrams to the economic evaluation of two medical interventions. Our analysis of pediatric cochlear implantation in Spain has proved that it is cost-effective with respect unilateral cochlear implantation, and the model for colorectal cancer screening with immunochemical fecal occult blood test showed that it is cost-saving with respect to no screening.

List of Figures

List of Tables

List of Algorithms

List of Acronyms

AI Artificial Intelligence

BCI Bilateral Cochlear Implantation

BN Bayesian Network

CEA Cost-Effectiveness Analysis

CEP Cost-effectiveness Partition

CI Cochlear Implant

CRC Colorectal Cancer

DAN Decision Analysis Network

DSA Deterministic Sensitivity Analysis

DSD Decomposition into Symmetric DANs

DT Decision Tree

EBUS Endobronchial ultrasound

EUS Transesophageal ultrasound-guided fine needle aspiration

EVPI Expected Value of Perfect Information

FDA Food and Drug Administration

FOBT Fecal Occult Blood Test

gFOBT guaiac FOBT

HCC Half-Cycle Correction

ICER Incremental Cost-Effectiveness Ratio

ID Influence Diagram

iFOBT immunochemical FOBT

INE Spanish National Institute of Statistics

LT Life Table

MED Mediastinoscopy

MID Markov Influence Diagram

NMB Net Monetary Benefit

PGM Probabilistic Graphical Model

PSA Probabilistic Sensitivity Analysis

QALE Quality-Adjusted Life Expectancy

QALYs Quality-Adjusted Life Years

RCTs Randomized Control Trials

TNM Tumor-Node-Metastases

TTO Time Trade-Off

VAS Visual Analog Scales

WHO World Health Organization

WTP Willingness To Pay

Part I

Introduction

1

Introduction

"Research is what I'm doing when I don't know what I'm doing."

Wernher von Braun Read

1.1 Motivation

Decision support systems help decision makers to allocate a finite set of resources efficiently. Given that each decision maker may have different effectiveness criteria and different risk aversion, it is important to establish objective methodologies that can support those decisions in order to improve the use of finite resources. It is also necessary to have ways of representing and solving complex problems. None of these models can represent all the complexity of the real world but they may offer a useful simplification that can explain and solve those problems. Artificial intelligence (AI) includes multiple useful frameworks to work with complex problems. In this thesis I focused on a particular AI field, the knowledge representation and reasoning. This field tries to represent real world problems and to perform complex tasks over them, such as Cost-Effectiveness Analysis (CEA). One of the frameworks in this field are the Probabilistic Graphical Models (PGMs).

PGMs have proved to be a good way to represent and analyze complex problems. PGMs have an intuitive and compact data representation and allow efficient reasoning using general-purpose algorithms. They can also be applied to several domains such as medical diagnosis, natural language processing, traffic analysis, social network models, computer vision, planning...

The main goal of the Research Center for Intelligent Decision-Support Systems (CISIAD) at the Universidad Nacional de Educación a Distancia (UNED), is to do research on PGMs and their application in diagnosis and decision making, especially in medicine. PGMs can encode complex temporal Markov models, which are the most popular modeling framework in health technology assessment, and perform cost-effectiveness and sensitivity analysis, which are the most common methods for economic evaluation.

1.2 Objectives

The aim of this thesis is to perform CEA with new AI models. To achieve this goal, I elaborated a research plan in which I considered three main tasks related to developing and applying algorithms for performing CEA on PGMs. However, during

the evaluation of existing medical models, we observed a common issue in the standard evaluation methodologies when the models have discontinuities. This new research line was introduced as a fourth objective.

1. Developing new algorithms for cost-effectiveness and sensitivity analysis with Markov Influence Diagrams (MIDs) (Díez et al., 2017).
2. Developing new algorithms for CEA with Decision Analysis Networks (DANs) (Díez et al., 2018a) to solve large asymmetric problems such as models in which the decisions are not totally ordered.
3. Applying PGMs to two real medical decision problems:
   a) Cost-effectiveness of pediatric Bilateral Cochlear Implantation (BCI) in Spain.
   b) Cost-effectiveness of ColoRectal Cancer (CRC) screening with the fecal occult blood test in Spain.

   Although these analyses were done for Spain, the models built can be applied in other countries with minor modifications.
4. Evaluating Markov models with within-cycle corrections (adjustments made to reduce the error introduced when the time is discretized) when the model has discontinuities.

1.3 Methodology

First of all, I reviewed the state of the art about PGMs and health economics evaluation.

The next step was working on the development and improvement of the algorithms for cost-effectiveness and sensitivity analysis on MIDs, a new type of PGM proposed by the CISIAD (Díez et al., 2017). Even though MIDs were already defined when I started this thesis, small modifications, such as the definition of criteria in utility nodes, were needed to perform CEA. Then we applied these algorithms to two medical problems: CRC screening and pediatric BCI.

For each of these medical applications we had to study the state of the art. We then selected as basis the most relevant models in the literature and tried to replicate

them in OpenMarkov . This process helped us to determine the relevant variables and causal relationships in the model. Taking into account all this information and the models in the literature, we built new models following an iterative approach: defining each block of the model, building and testing it to ensure that the underlying logic of the model was correct. The next steps in the modeling process were ruled by two simple criteria: the first one was to detail as much as possible the real problem (given the current data) and the second one to search for missing data that could give more detail to our model. When the assumptions taken in the model were not relevant to the expected results of the CEA (taking into account the uncertainty over the parameters), we entered a phase of testing—we had tested the functionality of many parts of each model several times during the construction, but in this phase, we tested the model as a whole. In this final phase we also obtained feedback from experts that helped us to improve our models.

To accomplish the implementation of CEA on DANs, we previously designed the algorithms based on the ones we developed for MIDs. DANs were already defined when I started this thesis, but it was necessary to adapt the existing algorithms to perform CEA. These algorithms and their adaptations were implemented in OpenMarkov, expanding their functionality and adapting each part to the capabilities of DANs. When the design and implementation were ready, we check these algorithms using unitary and functional tests.

The analysis of discontinuities in Markov models started during my short stay at the School of Health and Related Research (SCHaRR) of the University of Sheffield, where I was able to work with some of their models. As stated in the previous section, during the evaluation of a medical decision model we found a problem with some mid-cycle adjustment techniques when the models have discontinuities. After reviewing the state of the art and analyzing the different mid-cycle adjustment techniques proposed in the literature, we made an empirical study in which we compared each of these techniques with a gold standard. Even though the results obtained in the empirical approach offered some useful conclusions, we made the decision that it was necessary to formalize the theoretical demonstration in order to be able to generalize these conclusions.

1.4 Organization of the thesis

This document has been structured in five main parts:

- Part I explains the motivation, objectives, methodology, and organization of this thesis.
- Part II reviews the state of the art about PGMs and medical decision making.
- Part III presents the main methodological contributions of this thesis, describing the cost-effectiveness algorithms developed for MIDs and DANs and our analysis of within-cycle corrections when applied to Markov models with discontinuities.
- Part IV describes the two decision support systems developed to analyze the cost-effectiveness of pediatric BCI and CRC screening.
- Finally, Part V, presents the conclusions and discuss some research lines for future work.

Part II

State of the art

2

Probabilistic graphical models

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

2.1 Introduction

The first known treatise on probability, "Liber de ludo aleae", attributed to Gerolamo Cardano, was written in the 1560s but was not published until 1663. Around 1654, Fermat, developed with Blaise Pascal the principles of the theory of probability, as we know it today, while they tried to solve some problems posed by the also mathematician and player-Chevalier de Méré. Those writings were formalized by Christiaan Huygens in his treaty of "De Ratiociniis in Ludo Aleae" Calculating in Games of Chance) in 1657, which was the first published work on probability calculation. In 1763, two years after its author's death, it was published the work of Thomas Bayes published "Essay Towards Solving a Problem in the Doctrine of Chances", which serves as the basis for the theorem that carry his name. This theorem states the principles that allow us to use a model that tell us the conditional probability of an event A given the event B.

(Euler, 1741) is considered as the first article on graph theory. This article solved the problem of the Köningsberg Bridges over the river Pregel, which consists in find a path that goes through the seven bridges without crossing the same bridge twice. In such theories only discrete probability spaces were involved and the analysis methods were solely combinational.

The joint use of graphs and probabilistic distributions is applied in a wide spectrum of knowledge areas. In engineering, this idea was first developed by Kirchhoff (1847) to analyze electrical networks. In statistical physics, Gibbs (1902) used an undirected graph to represent the distribution over a system of interacting particles. In 1921 the graphical idea applied to Bayes problems was developed by Wright (1921) in his paper "Correlation and causation". Wright was one of the pioneers in the use of graphical method and proposed the use of a directed graph to study inheritance in natural species, a method which is still broadly used.

In 1933, Kolmogorov (1933) published the book on the fundamentals on probability theory establishing the foundations of modern probability theory. Bartlett (1935) proposed the idea of analyzing interactions between variables. The first book on graph theory was written by König (1936)—see (Tutte, 2001). In the field of computer science, Turing (1950) presented methods for building computerized expert systems designed to carry out difficult tasks, such as solving to complex problems or performing

medical diagnosis. Some early AI systems used probabilistic methods based on the very restrictive naive Bayes model (Lodwick et al., 1963; Warner et al., 1964; de Dombal et al., 1972; Gorry, 1973). The idea was also adopted in economy and social science; e.g. Wold (1954) used it in his is article "Causality and econometrics" and Blalock (1971) in his book "Causal Models in the Social Sciences", both authors showed the usefulness of this tool in their respective fields.

Although the naive Bayes model restricts itself to a small set of possible hypotheses (e.g., diseases) and assumes that the findings (e.g., symptoms or test results) are conditionally independent given each hypothesis, it was surprisingly successful, performing (within its area of expertise) at a level comparable to or better than that of experts. For example, the system of de Dombal et al. (1972) averaged over 91.8 percent correct diagnoses of acute abdominal pain, whereas expert physicians were averaging around 79.6 percent. Nevertheless, much uncertainty and confusion remained about the role of probability in AI (Szolovits and Pauker, 1978).

In the 1980s the probabilistic methods began to have a major acceptance, driven forward by two main factors. The first was the theoretical development of Bayesian Networks (BNs) by Judea Pearl (1986; 1988). Simultaneously the paper by Lauritzen and Spiegelhalter (1988) proposed an efficient algorithm for PGMs. The second important factor was the construction of large-scale, highly successful expert systems based on this framework. Two of the most visible of these applications were the Pathfinder, an expert system for hematopathology diagnosis (Heckerman et al., 1992) and the MUNIN, a causal probabilistic network for interpretation of electromyographic findings (Andreassen et al., 1987).

Nowadays, probabilistic methods in general, and PGMs in particular, have gained almost universal acceptance in a wide range of disciplines and fields, generally in processes of modeling, simulation, and knowledge representation.

2.2 Basic definitions

2.2.1 Graphs

A graph (G) consists of a set of nodes (V) and edges (E). An edge is a binary relation between two nodes (X, Y), where $X, Y \in V$. If X and Y are ordered in their relation, the edge is directed; if not, the edge is undirected. Directed edges are also known as arcs or directed links and can be represented by an arrow; e.g. if the edge (X, Y) is directed we can represent it as the arc $X \to Y$. We say that a graph is directed when all its edges are directed and, it is undirected when all its edges are undirected.

When there is an arc $X \to Y$, we can say that X is the parent of Y and that Y is the child of X. The set of parents of a node X is denoted by $Pa(X)$ and the set of its children by $Ch(X)$.

A path is an ordered set of distinct nodes $\{X_0, X_1, ..., X_n\}$ and edges such that there is an edge that links every pair of adjacent nodes; i.e. there is an edge (X_0, X_1), an edge (X_1, X_2), etc. The path is directed if all the edges that conform the path are directed and are oriented in the same direction. All the nodes for which there is a directed path to X are known as the ancestors of X and are denoted by $An(X)$. The nodes for which there is a directed path from X are known as the descendants of X and are denoted by $De(X)$. In graphs used for PGMs there can be only one link between each pair of nodes.

A cycle in a directed graph is a path $\{X_0, X_1, ..., X_n\}$ in which the last and first node are linked by an edge; i.e. $(X_0 \to X_n)$. A directed graph that does not have any cycle is called acyclic directed graph.

A tree is an acyclic directed graph in which each node has only one parent, except the root node, which has no parents. The nodes having no children are known as leaves.

2.2.2 Probabilistic graphical models

A PGM consist of a graph $G = (V, E)$ and a probability distribution P over a set of variables $\boldsymbol{X}$, such that each node in V represent one of the variables. As every node in the graph represents one of the variables of the PGM, we are going to use the terms node and variable indifferently. The graph can only contain one link between each pair of nodes. There are several types of PGMs: some of them use acyclic directed graphs (for example BNs, influence diagrams, etc.), others use undirected graphs (Markov networks, also known as Markov random fields), while others use acyclic partially directed graphs (Koller and Friedman, 2009; Sucar, 2015). The absence of an edge in the graph represents a relation of independence in P. Therefore there is a relationship between the graph and the probability distribution, which determines how probability can be factored.

2.2.3 Bayesian networks

A BN (Pearl, 1988) is a PGM whose graph is directed and acyclic. Unlike decision models contains only one type of node. The probability distribution can be factored as

$$P(\boldsymbol{x}) = \prod_i P(x_i \mid pa(X_i)) \,, \tag{2.1}$$

where $pa(X_i)$ denotes a configuration of the parents of X_i, i.e. an assignment of values to each one of the parents of X_i; and $P(x_i|pa(X_i))$ are the conditional probabilities obtained from $P(\boldsymbol{x})$. This equation involves a conditional probability distribution for each node in the graph.

BNs satisfy two properties which also can be used as alternative definitions:

1. Markov property. Each node is independent of its non-descendants given its parents, i.e., if Y is a set of nodes such that none of them is a descendant of X, then $P(x|Pa(X), y) = P(x|Pa(X))$.
2. d-separation. Two nodes X and Y are d-separated in the graph given a set of nodes $\boldsymbol{Z}$, when there is no active path connecting them. A path is active if every node W between X and Y satisfies this property: if the arrows that connect W

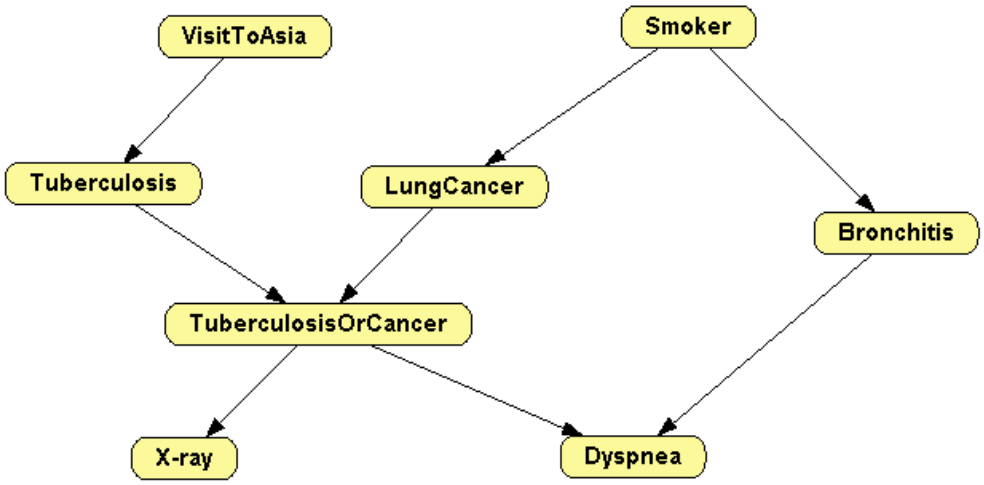


Figure 2.1: BN for the diagnosis of the dyspnea causes (Lauritzen and Spiegelhalter, 1988).

with its two neighbors converge in it, then W or at least one of its descendants is in $\boldsymbol{Z}$; else, W is not in $\boldsymbol{Z}$. If X and Y are d-separated in the graph given a set of nodes $\boldsymbol{Z}$, which we denote by $I_G(X, Y|\boldsymbol{Z})$, then they are probabilistically independent given $\boldsymbol{Z}$: $\forall X, \forall Y, \forall \boldsymbol{Z}, I_G(X, Y|\boldsymbol{Z}) = I_P(X, Y|\boldsymbol{Z})$.

We will assume in this thesis that BNs are discrete, which means that each variable has a finite set of states; i.e. the domain of a variable X_i is $dom(X_i) = (x_i^1, x_i^2, ..., x_i^n)$. Figure 2.1 shows an example of a BN.

In practice it may be useful to know the a posteriori probability of some variable of interest given a set of findings. A finding is the determination of the value of a variable, $X_i = x_i$, from the observed data; e.g. determining that a patient is a smoker. The evidence is the set of the all available findings at a given time or situation, i.e. the assignment of a value to each variable in set $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$.

2.3 Decision models

Decision theory is the study on how and why the decisions are made. This can be analyzed from two different perspectives. The first one, called "descriptive", studies the rationale underlying a decision and tries to explain the process that decision

makers follow. The second perspective, which is the one we are interested in, is known as "normative" and describes a complete framework to help decision makers. This approach consists in giving the decision maker useful information in order to help to reduce the uncertainty of the expected results before making one or another decision. Decision analysis includes the evaluation of each of the possible events or consequences of making a decision, but must also evaluate all the alternatives to that decision, taking into account the opportunity costs.

2.3.1 Decision trees

A decision tree (DT) (Raiffa, 1968) is a tree having three types of nodes: chance, decision and value. Chance nodes, drawn as circles, represent events which are not under the direct control of the decision maker. Decision nodes, drawn as rectangles, correspond to actions under the direct control of the decision maker. Decision nodes represent all the alternatives that the decision maker has. This set of possible decision rules must be exhaustive and exclusive; i.e. all the alternatives must be represented and they must be incompatible between them. Value nodes, drawn as triangles or hexagons, are the leaves of the DT and each one represents the expected outcomes of a particular scenario. Figure 2.2 shows an example of a DT.

Each node in the DT (except the leaves) represents a chance or decision node. We build a DT from its root node to its leaves. Each of the branches (links) of a chance node are labeled with one of the possible states of the variable and the conditional probability of that state, while the branches of decision nodes represents the available options. The leaves of the DT are the expected outcomes of each possible scenario; i.e. each leaf has a value conditioned on the configuration of all the nodes in its path, from the root to the leaf.

The evaluation of a DT (Raiffa, 1968) is computed from the leaves to the root. When evaluating a chance node we calculate the weighted sum of its branches; i.e. the sum of the values of each branch weighted by its probability:

$$U_X(Pa(x)) = \sum_X U(x|Pa(x)) \cdot P(x|Pa(x)) \tag{2.2}$$

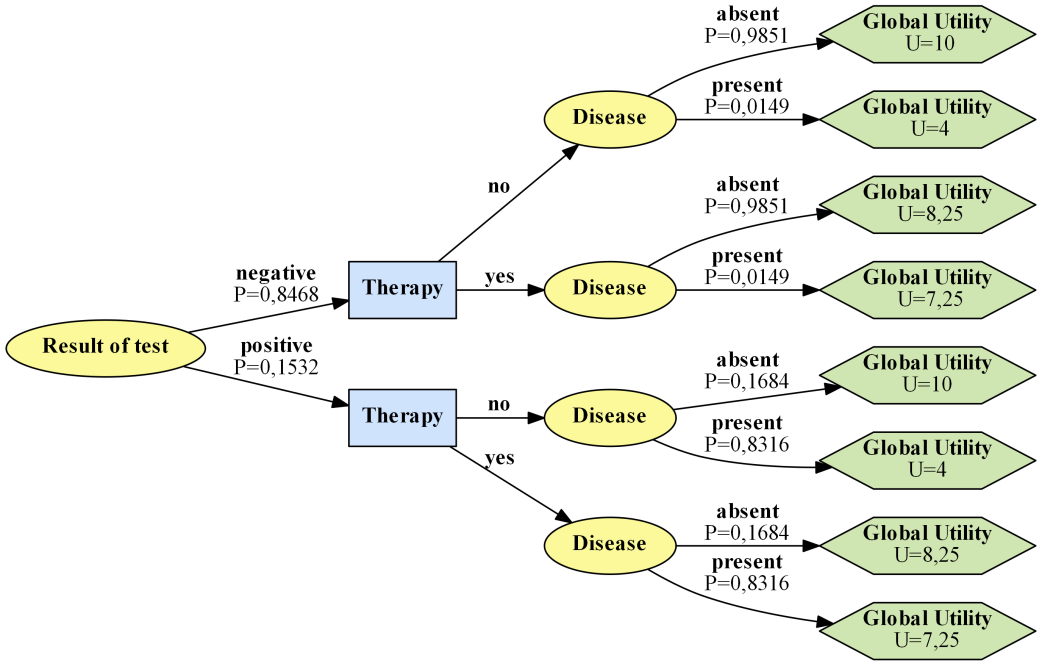


Figure 2.2: Graphical representation of a DT.

When evaluating a decision node we must to get the decision that give us higher utility; the optimal policy for this decision that is obtained as the maximum utility between its branches:

$$U_D(Pa(d)) = \max_d U(d|Pa(x)) \tag{2.3}$$

In Figure 2.3 we can see the previous example solved. The expected outcome of the disease in the scenario in which the result of the test is negative no therapy is applied is 9.9107, that is obtained from:

$$\begin{aligned} U_{Disease}(scenario) &= \sum_{Disease} U(disease|scenario) \cdot P(disease|scenario) \\ &= U(disease = no|scenario) \cdot P(disease = no|scenario) \\ &+ U(disease = yes|scenario) \cdot P(disease = yes|scenario) \\ &= 10 \cdot 0.9851 + 4 \cdot 0.0149 = 9.9107 \end{aligned}$$

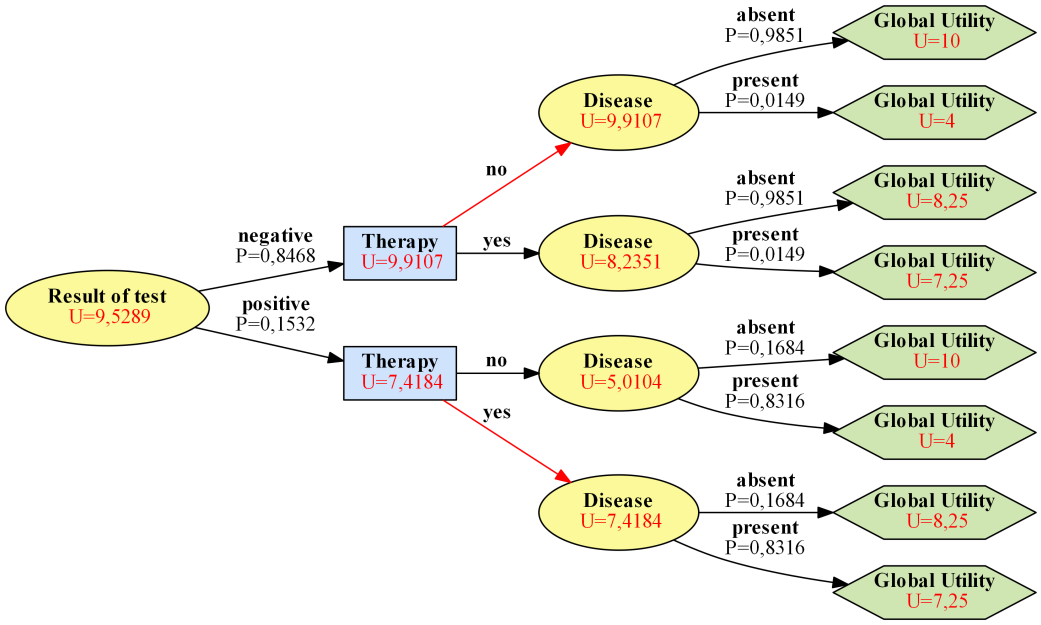


Figure 2.3: Evaluation of a DT.

When the consequences of the interventions are not deterministic, it is necessary to model the probability of each outcome. Algorithm 2.1 shows the standard roll-back algorithm for the evaluation of unicriterion DTs.

Algorithm 2.1: Roll-back algorithm for unicriterion DTs.

Input: A decision tree

Result: The expected utility and a policy for each decision node

1 **foreach** *node* n **do**
2 **if** n *is a chance node* **then**
3 $u_n = \sum_i p_i \cdot u_i$, where p_i is the probability of the i-th branch and u_i is its utility
4 **if** n *is a decision node* **then**
5 $u_n = \max u_i$

DTs are the tool used most frequently for decision analysis, especially in medicine (Pauker and Wong, 2005). They have the advantage of almost absolute flexibility, but also have four drawbacks.

First, their size grows exponentially with the number of variables. In general a

variable gives rise to several nodes in the DT.

Second, a DT cannot represent conditional independencies; so the graph of the tree does not represent the structure of the problem nor the hypotheses implicit in the model.

Third, DTs require in general a preprocessing of the probabilities (Howard and Matheson, 1984a; Bielza et al., 2011); for example, medical diagnosis problems are usually stated in terms of direct probabilities, namely the prevalence of the diseases and the sensitivity and specificity of the tests, while DTs are built with inverse probabilities, i.e., the positive and negative predictive values of the tests. Even in cases with only a few chance variables, this preprocessing of probabilities is a difficult task. For these reasons, DTs can only represent small problems. In the medical literature, trees usually have 3 or 4 variables and between 6 and 10 leaf nodes. A tree of 5 variables typically contains around 20 leaf nodes, which implies that building, debugging, and analyzing it would require a significant effort. Given that the maximum size of DTs in practice is of the order of 50 nodes, they can only solve problems of at most 6 or 7 variables.

And fourth, standard evaluation algorithms cannot perform CEA when the tree contains embedded decision nodes, i.e., nodes that are not the root of the tree. Even worse, TreeAge, which is by far the most common software package for DTs, may return wrong results in this case, without warning the decision analyst (Arias and Díez, 2014).

2.3.2 Influence diagrams

Influence diagrams (IDs) (Howard and Matheson, 1984a; Pauker and Wong, 2005), in contrast to DTs, have the advantages of being very compact, representing conditional independencies, and using direct probabilities, i.e., the probability of the effect conditioned on the cause.

IDs are a generalization of BNs used to model and solve decision problems. An ID consists of an acyclic directed graph $G = (V, E)$, where the set V has three types of nodes: chance nodes $\boldsymbol{V}_C$, decision nodes $\boldsymbol{V}_D$, and value nodes $\boldsymbol{V}_U$, as shown in Figure 2.4.

Figure 2.4: Types of nodes in an ID.

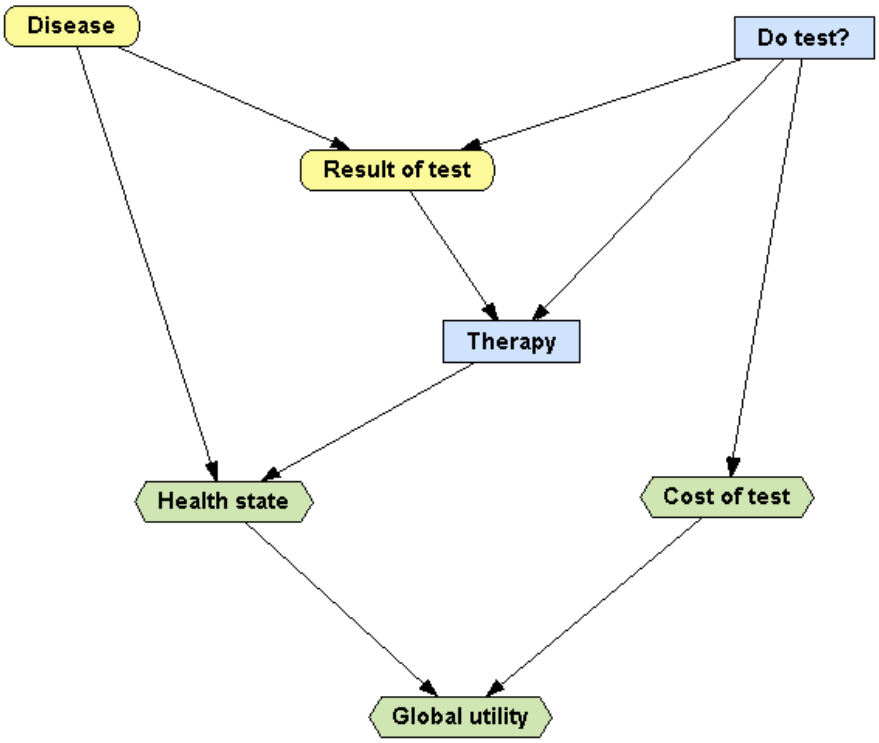


Figure 2.5: Types of arcs in an ID. Figure extracted from (Arias and Díez, 2014).

In IDs, each chance node $X_i \in \boldsymbol{V}_C$ represents a random variable with a conditional probability distribution while each decision node $D_i \in \boldsymbol{V}_D$ represents a controllable variable whose value is determined by the decision maker with a decision rule (also known as policy), $\delta_i : Pa(D_i) \rightarrow D_i$, which determines the values of d_i based on the observed values. The collection of policies constitutes an strategy.

Value nodes, $\boldsymbol{V}_U$, drawn as diamonds or hexagons, represent the expected benefit or loss, or more generally, the outcome for the decision maker, in other words value nodes represent functions that the decision maker wants to maximize. Value nodes cannot be parents of chance or decision nodes.

Depending on the type of node they point at, there are three types of links or arcs in an ID. Links pointing at chance nodes represent probabilistic dependencies (as

in BNs); for example, the link *Disease* $\rightarrow$ *Result of test* in Figure 2.5. Links pointing at decision nodes, such as *Result of test* $\rightarrow$ *Therapy*, are named information arcs and represent availability of information. Thus like $X \rightarrow D$ from a chance node X to a decision node D, means that the state of X is known before decision D is made. Finally, links pointing at value nodes represent functional dependence, such as the link *Disease* $\rightarrow$ *Health state*.

It is assumed that and ID has no barren nodes—a chance node or a decision node that does not precede any other node, or that all its descendants are barren—since they have no impact on the decisions (Nielsen and Jensen, 1999).

We assume that there is a a total ordering of the decision nodes, given by a directed path that connects all the decision nodes.

It follows from this definition that in an ID with a set of n decisions $D_1, ..., D_n$, the set of informational arcs induces a partitioning of chance variables $\boldsymbol{V}_C$ into $n+1$ disjoint subsets, $\{\boldsymbol{C}_0, ..., \boldsymbol{C}_n\}$, where $\boldsymbol{C}_i$ contains every chance variable C with an arc $C \rightarrow D_i$ but with no arc $C \rightarrow D_j$ to a previous decision, $j < i$; i.e., $\boldsymbol{C}_i$ is the subset of chance variables known for D_i but unknown for previous decisions. This induces a partial order $\prec$ in $\boldsymbol{V}_C \cup \boldsymbol{V}_D$:

$$\boldsymbol{C}_0 \prec \{D_0\} \prec \boldsymbol{C}_1 \prec ... \prec \{D_n\} < \boldsymbol{C}_n \tag{2.4}$$

The set of variables whose value is known by the decision maker when making the decision D_j are called the informational predecessors of D_j and are denoted as $iPred(D_j)$. The no-forgetting property of IDs, also known as perfect recall assumption (Koller and Friedman, 2009), states that the decision maker remembers all previous decisions and observations. It implies that $iPred(D_i) \subseteq iPred(D_j)\ \forall i \leq j$. In particular, $iPred(D_j)$ is the set of variables that occur before D_j, i.e. $iPred(D_j) = \boldsymbol{C}_0 \bigcup D_0 \bigcup \boldsymbol{C}_1 \bigcup ... \bigcup D_{i-1} \bigcup \boldsymbol{C}_i$.

The quantitative information that defines an ID is given by assigning to each probability node C a probability distribution $P(c \mid pa(C))$ for each configuration of its parents (as in BNs), and assigning to each value node U a function $\psi_U(pa(U))$ that maps each configuration of its parents onto a real number.

The domain of each function ψ_U is given by its functional predecessors $fPred(U)$. In the case of an ordinary utility node, the functional predecessors will be the parents of the utility node.

For each configuration $\mathbf{v_D}$ of the decision variables V_D, we have a probability distribution over the chance variables $\mathbf{v_C}$:

$$P(\mathbf{v}_\mathcal{C} : \mathbf{v}_\mathcal{D}) = \prod_{C \in \mathbf{V}_\mathcal{C}} P(c \mid Pa(C)) \ , \tag{2.5}$$

that represents the probability of configuration $\mathbf{v_C}$ when the decision variables are set to the values $\mathbf{v}_\mathcal{D}$.

A stochastic policy for a decision D is a probability distribution defined over D and conditioned on the set of its informational predecessors, $P_D(d \mid iPred(D))$. If P_D is degenerate, i.e. it only has zeros and ones, we can say that the policy is deterministic. A deterministic policy could be understood as a function π_D that assigns to each configuration of $iPred(D)$ a value d of D.

A strategy Δ is a set of policies, one for each decision, $\{P_D \mid D \in \mathbf{V}_D\}$. A strategy Δ induces a joint distribution over $\boldsymbol{V}_C \cup \boldsymbol{V}_D$ defined as follows:

$$\begin{aligned} P_\Delta(\mathbf{v}_C, \mathbf{v}_\mathcal{D}) &= P(\mathbf{v}_\mathcal{C} : \mathbf{v}_\mathcal{D}) \prod_{D \in \mathbf{V}_\mathcal{D}} P_D(d \mid iPred(D)) \\ &= \prod_{C \in \mathbf{V}_\mathcal{C}} P(c \mid Pa(C)) \prod_{D \in \mathbf{V}_\mathcal{D}} P_D(d \mid Pa(D)) \end{aligned} \tag{2.6}$$

The expected value of a strategy Δ is defined as:

$$EU(\Delta) = \sum_{\mathbf{v}_\mathcal{C}} \sum_{\mathbf{v}_\mathcal{D}} P_\Delta(\mathbf{v}_\mathcal{C}, \mathbf{v}_\mathcal{D}) \psi(\mathbf{v}_\mathcal{C}, \mathbf{v}_\mathcal{D}) \, , \tag{2.7}$$

where ψ is the value associated with the node U_0, that is the global outcome. The optimal strategy is one that allows obtaining the maximum expected value:

$$\Delta_{opt} = arg \max_{\Delta \in \Delta^*} EU(\Delta) \ , \tag{2.8}$$

where Δ^* is the set of all strategies for the ID.

The expected value of an ID, also known as maximum expected value, is the one obtained when applying the optimal strategy:

$$EU = EU(\Delta_{opt}) = \max_{\Delta \in \Delta^*} EU(\Delta) \tag{2.9}$$

The evaluation of an ID consists in find the maximum expected value and the optimal strategy. It can be performed by applying this equation:

$$EU = \sum_{\mathbf{c_0}} \max_{d_0} ... \sum_{\mathbf{c_{n-1}}} \max_{d_{n-1}} \sum_{\mathbf{c_n}} P(\mathbf{v}_\mathcal{C} : \mathbf{v}_\mathcal{D}) \psi_{U_0}(\mathbf{v}_\mathcal{C}, \mathbf{v}_\mathcal{D}) \,. \tag{2.10}$$

The optimal strategy of an ID always can be obtained and, in case that more than one strategy gives the maximum expected utility, we will say that there exist multiple optimal strategies.

IDs can be evaluated by obtaining their equivalent DTs. The first step to build an equivalent DT is to determine the order in which the variables will appear on the tree. There are two main rules to define this order:

- If we have two decisions D_1 and D_2 and D_1 is made before D_2, then D_1 must be placed at the left of D_2. Remember that for IDs we must know the total ordering of the decisions.
- The variables that are known before taking a decision D_1; i.e. the informational predecessors of the decision $iPred(D_1)$, must appear at the left of that decision. The other variables (those unknown before making a decision) must appear to the right of the decision.

Applying the rules above we can say that, when a variable X is known before making a decision D_2 but it is unknown before making a decision D_1, that variable will be placed between that decisions in the tree. Once the structure of the DT is defined, we must indicate what are the expected outcomes for each of the leaves of the tree.

There are several evaluation algorithms (Shachter, 1986; Jensen et al., 1994; Dechter, 1996; Cowell et al., 1999) that can evaluate IDs. During this thesis we used variable elimination(Shachter, 1986).

Variable elimination evaluate the variables one by one according to a certain

order (that can be defined with an heuristic). Variables that have been evaluated can be removed from the model, and this process continues until all the variables have been evaluated.

To eliminate a variable it is necessary to perform a series of transformations on the ID; these transformations are guaranteed to preserve the optimal policies and the expected outcomes. The operations that can be made to the network are:

- Eliminate barren nodes, chance or decision nodes that are leaves in the graph; i.e. the nodes that do not affect the decisions or expected outcomes.
- Eliminate chance nodes whose all children are value nodes and do not have other children. In this case, the outcome is updated according to the result of marginalizing in this variable the joint probability function.
- Eliminate decision nodes whose children are value nodes and which all its ancestors are also parents of those value nodes. The evaluation of this decision nodes and its children (the value nodes) result on a potential that maximizes the expected outcome. The policy that maximizes the expected outcome is added to the optimal policies set and the outcomes function is updated.
- If none of the previous operations can be applied, we must apply the arc-reversal algorithm between chance variables. It consists in transforming a decision or chance node, so that the transformation puts us in one of the situations described above (transforming it in a barren node, or in nodes whose children be only value ones). To make arc-reversal between nodes X_i and X_j it is required that there be no other trajectory between these nodes. Then the arc $X_i \to X_j$ is inverted and each node inherits the parents of the other node.

2.3.3 Markov influence diagrams

MIDs (Díez et al., 2017) are PGMs that extend IDs in the same way as Markov decision trees extend DTs (Beck and Pauker, 1983; Sonnenberg and Beck, 1993)—allowing the representation of repetitive events, as we will see in this section. MIDs allow building state-transition models. As IDs, MIDs consist of a graph—that defines the relationships between the variables (generally causal)—and a set of potentials that

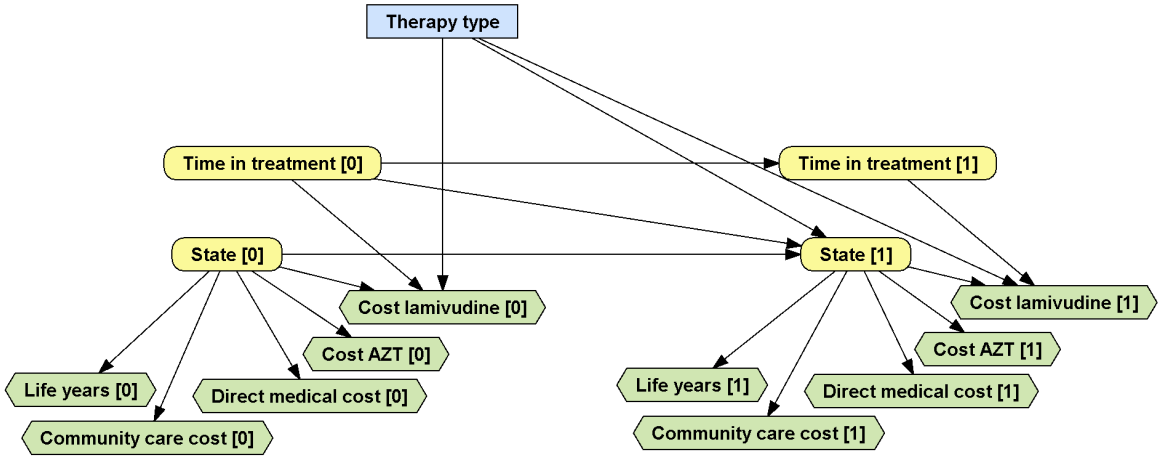


Figure 2.6: MID for the HIV model of Chancellor et al. (1997).

include the probability distributions and the utility functions. Figure 2.6 shows an example of an MID.

They are especially designed to perform CEA. Using a causal graph that may contain several variables per cycle, MIDs can model various features of the patient without multiplying the number of states; in particular, they can represent the history of the patient without using tunnel states (sub-states that can be visited only in a fixed sequence, normally used to represent temporal permanence on a state).

The main novelty presented by MIDs versus IDs is that they allow representing variables that evolve over time. These temporal variables are represented by an index in brackets, which express the cycle to which the node belongs, for example *State [0]* in Figure 2.6 belongs to the cycle *0*. MIDs also allow representing non-temporal variables such as *Therapy type*. In this way, if we take the set of nodes that belongs to the cycle i and its arcs, we can talk about the temporary slice in the i-th cycle. MIDs can represent the temporal evolution of chance variables and utility variables of the model.

An MID can be expanded by representing all the nodes and links in each of cycles.

In Figure 2.6 we can observe the MID representation of the model of Chancellor et al. (1997). This model compares two possible treatments for HIV. The link *State[0]* $\rightarrow$ *State[1]* implies that the health state of a patient depends on his/her health state in the previous cycle; i.e. the state in which it was in the previous cycle, and of the time

that the patient was in treatment. In this model, the cycle length is one year and the temporal horizon is 20 years. In Figure 2.6 we can see the expanded network for two cycles. The network can be expanded to the number of desired cycles to evaluate a specific time horizon. This expansion will replicate the slice of the last period to create the nodes and relationships of the subsequent periods.

Due to the computational cost, MIDs containing temporal decisions can only be evaluated for very short horizons. For this reason we will assume that all decisions are atemporal, i.e., they are made at the beginning of the process; they may be conditioned on future events (for example, "do the test when the symptom is present"), but the policy is the same for all cycles. This way it is possible to evaluate MIDs with large horizons. This is the main difference between MIDs and dynamic IDs (Tatman and Shachter, 1990), and leads to a completely different evaluation of the model. To our knowledge, the restriction to atemporal decisions is also present in all economic evaluation models, including cohort models, patient-level simulation and discrete event simulation. An important difference, however, is that cohort models usually represented the patient's state with a single variable, taking on a limited number of values (states), and consequently often need to multiply the number of states—for example, to represent the patient history—while patient-level simulation, discrete event simulation and MIDs can use several variables to model different features of the health state.

Like in IDs, there is a need for a directed path that goes through all the decision nodes, establishing the order in which the decisions are made. The only exception to this rule is when the order of the decisions is not relevant for the evaluation of the model (Nielsen and Jensen, 1999); i.e. when the order in which the decisions are made does not affect the maximum expected utility.

A temporal model is stationary after i cycles when all its potentials do not vary after the i-th cycle and no other $k' < k$ makes the model stationary. This property can be used to define only the first non-stationary i cycles and expand the network from the cycle i. Since the model is stationary in i it is enough to replicate the slice of that cycle the number of times necessary to reach the temporal horizon. The representation of the stationary model is known as compact representation of the model, while the representation of all the temporal horizon is known as expanded representation of the

model.

MIDs may contain temporal numerical variables, for example the value of the variable *Time in treatment* in Figure 2.6 increases by one in each cycle; i.e. it is 0 years in the first cycle, 1 year in the second cycle, etc.

An MID can be converted into an ordinary ID by expanding it to the desired horizon and then evaluated with the standard algorithms for IDs. Given that an can be converted into an ID, all the evaluation evaluation algorithms for IDs works also for processed MIDs.

2.3.4 Decision analysis networks

There are three types of asymmetries: order asymmetry, domain asymmetry and information asymmetry. Order asymmetry occurs when there are multiple decisions that can be made in different orders; we say that there is domain asymmetry when the value taken by a variable restricts the values that other variables can take and the information asymmetry occurs when a variable is observed only for some values of another variable (Jensen et al., 2006; Bielza et al., 2011; Díez et al., 2018a).

Several formalisms have been proposed for representing and solving asymmetric decision problems, but all of them have drawbacks; for example, unconstrained IDs (Jensen and Vomlelová, 2002) cannot represent domain asymmetry nor information asymmetry, and sequential IDs (Jensen et al., 2006) may need redundant links with complex labels.

DANs (Díez et al., 2018a) are a new type of PGM that allows to represent asymmetric problems. That is important because almost all the real-world problems are asymmetric.

Like IDs and MIDs, the graph of a DAN has three types of nodes: chance, decision and value, and each variable in a DAN has a potential.

DANs represent domain asymmetry by means of restrictions (Díez et al., 2018a). Each restriction is associated to a link, $X \to Y$, such that X and Y are chance or decision variables. A restriction is a pair (x, y), where x is a value of X and y is a value

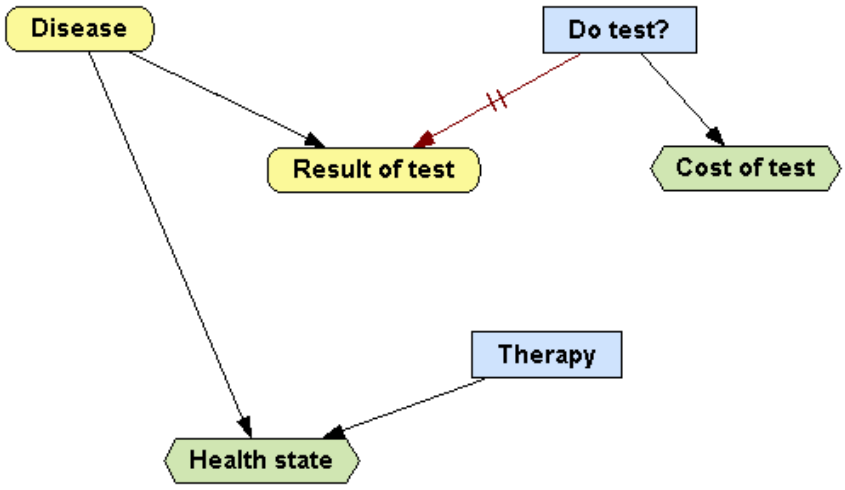


Figure 2.7: Example of a DAN.

of Y. It means that variable Y cannot take the value y when X takes the value x. In Figure 2.7 we can see a DAN in which the decision of not performing the test restricts the possible values of *Result of test*. The restriction in this figure is represented by a red arrow crossed by two parallel lines.

Information asymmetry is modeled with revelation links. A revelation link is an arc $X \rightarrow Y$, such that Y is a chance variable and X is a chance or decision variable, and certain values of X reveal the value of Y. The values of X that reveal the value of Y are known as *revelation conditions.* If all of the values of X reveal the value of Y we can say that X reveals Y unconditionally; otherwise we say that X reveals Y conditionally. All the conditional revelations induce information asymmetry, while unconditional revelations do not. Another way of representing the availability of information and satisfying the non-forgetting property (Howard and Matheson, 1984b; Nielsen and Jensen, 1999; Koller and Friedman, 2009) is by means of *always-observed* variables, whose value is always known before making any decision. Figure 2.8 shows an example of an always-observed variable, *Symptom,* marked with a red border, which is always known before making the decisions of performing the test and applying a therapy.

The decision in a DAN can be partially ordered. When decision D_1 is necessarily made before D_2we draw a link $D_1 \rightarrow D_2$. Please note that information links in IDs and MIDs are replaced with the revelation links in DANs. Every ID can be easily

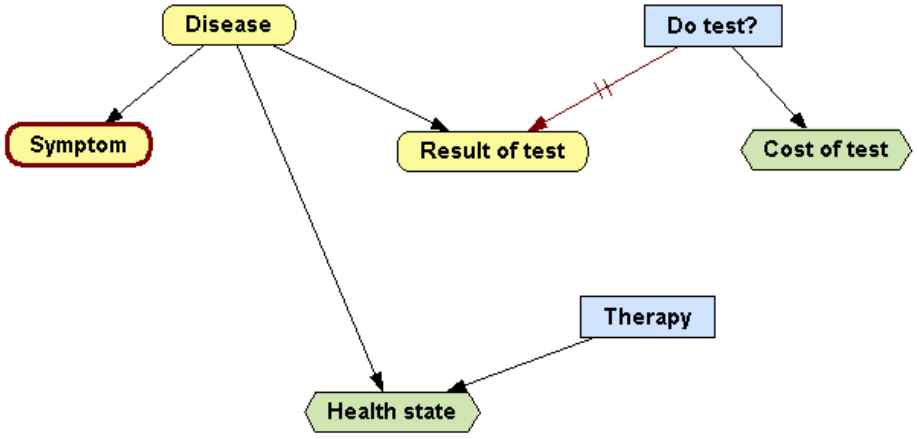


Figure 2.8: Example of an always-observed variable in a DAN.

transformed into an equivalent DAN, but for many DANs there is no equivalent ID. Therefore, IDs can be understood as a small subset of DANs.

The evaluation of DANs is described by Díez et al. (2018a) but basically there are two approaches: build an equivalent DT or decompose the DAN into symmetric DANs [1]. As the authors explain, symmetric DANs can be evaluated with the same algorithms as IDs, so it is possible to apply variable elimination algorithm to evaluate the symmetric DANs calculated by the Decomposition into Symmetric DANs (DSD) algorithm.

[1] A DAN is symmetric if it contains no asymmetry, i.e., it has no restrictions, if a value of a variable X reveals Y, then all the other values of X reveal Y, and a directed path connects all the decisions.

3

Medical decision analysis

"If we can make the correct diagnosis, the healing can begin. If we can't, both our personal health and our economy are doomed."

Andrew Weil

The main goal of a national healthcare system is to improve the health conditions of the population, trying that each one of its individuals have the best possible health state. Despite that, the resources of any health system are limited and insufficient to cover all medical needs of their population. Therefore it is important to make decisions about what treatments must be applied with that limited resources. For many years, decision making was based on subjective evaluations and patients could receive one or another therapy (or no one) depending on social, geographical, political or economic factors. Trying to establish objective criteria to make medical decisions and considering principles as the equality, several techniques have been developed to offer useful information for decision makers. To allow decision makers to have objective information to compare therapies, it is necessary to weigh the costs and outcomes of each alternative. Depending on whether we measure the health related outcomes we are really talking about one type of analysis. Cost-benefit analysis measures all health-related outcomes in economic units (Drummond et al., 2005). This approach was considered lacking in sensitivity, which has led into other measurements that represent both the effectiveness of the intervention and the benefit to the patient. In CEA health related outcomes are expressed in medical units, such as life expectancy, quality-adjusted life years or patient experience. In cost-utility analysis effectiveness is identified with Quality Adjusted Life Expectancy (QALE). This approach derived into the cost-utility analysis, a particular kind of CEA that measures the utility in Quality-Adjusted Life Years (QALYs).

3.1 Cost-effectiveness analysis

CEA is the most common framework in health economics (Siebert, 2003). This analysis compares the costs and effectiveness of different alternatives and order them according to their cost-effectiveness ratio. CEA is a particular type of multicriteria optimization problem with two criteria: cost and effectiveness. Given that a cost-utility analysis is by definition a CEA we are going to refer they also as CEA.

3.1.1 Quality of life

While direct and indirect costs derived from a treatment can be obtained objectively, the measurement of the quality of life is subjective. The estimation of quality of life is obtained from a group of informants. They can be patients suffering from a disease or people who are given all the information about the consequences of the disease. Therefore, it can vary depending on the demographic, sociological, historical, and personal factors of the informants. The more representative the sample of informants, the better it will represent the preferences of the general population.

The quality of life, QoL, is measured in a scale where 1 corresponds to perfect health and 0 to being dead. When the quality of life varies over time, the QALE corresponding to an interval of time $[t_1, t_2]$ is given by:

$$QALE = \int_{t_1}^{t_2} QoL(t) \cdot dt \tag{3.1}$$

There are different techniques for obtaining the quality of life associated with an intervention or treatment. The most common is the Time-Trade-Off (TTO) method, which consists in asking to each informant how much time he/she would give up in order to regain perfect health. For example, if one subject indicates that 10 years of life with a treatment equals 8.5 years of life with perfect health (that is, he would give up 1.5 years of life), we estimate that, for him, the quality of life associated with that treatment is 8.5 / 10 = 0.85.

3.1.2 Willingness to pay and net monetary benefit

The main purpose of CEA is to determine whether the effectiveness of an intervention outweighs its economical costs (Drummond et al., 2015; Neumann et al., 2016). One of the approaches is to convert our CEA into a cost-benefit problem—in which all the measures are converted to economic units. To transform the effectiveness, measured in QALYs, into economic units we must know the Willingness To Pay (WTP) of the decision maker, also known as cost-effectiveness threshold. The WTP is the amount, in economic units, that a decision maker is willing to pay in order to gain one

quality-adjusted life year. Given the WTP (λ) of a decision maker we can calculate the Net Monetary Benefit (NMB, Eq. 3.2) to know if an intervention is cost-effective for that decision maker. If the NMB is positive the effectiveness (E_A) outweighs the costs (C_A) of that intervention while if the NMB is negative the intervention is not cost-effective for this decision maker as the costs are greater than the effectiveness. When there are several interventions, the problem is determine which one maximizes the NMB (Stinnett and Mullahy, 1998).

$$NMB_A(\lambda) = \lambda \cdot E_A - C_A \tag{3.2}$$

The WTP is a subjective value that depends on each decision maker so it is impossible to determine whether an intervention is cost-effective or not without fixing a WTP value. The WTP of a national health care system depends on factors as the budget constraint and diverse social aspects. In Spain, given the lack of an official value, several researchers have tried to estimate the shadow threshold; the results obtained are €20,000/QALY (Pinto and Rodríguez, 2001), €30,000 per life year gained (Sacristán et al., 2002), €24,000-42,000/QALY (Soto, 2004), €9,000-38,000/QALY (Pinto and Martínez, 2005), and €30,000-45,000/QALY (de Cock et al., 2007). The threshold usually accepted as a consensus among experts and most commonly used in Spanish economic evaluations (Catalá-López et al., 2016) is €30,000/QALY.

In the UK, it has been accepted the WTP £20,000-£30,000/QALY (NICE, 2013) to decide whether an intervention should be applied. In the US the estimate of the WTP is $50,000-$100,000 based on the studies of Kaplan and Bush (1982) and Ubel et al. (2003).

Given that the WTP is a subjective measure, in the field of health economics different interventions are usually compared through the Incremental Cost-Effectiveness Ratio (ICER).

3.1.3 Deterministic cost-effectiveness analysis

CEA is a multi-objective problem in which we want to maximize the effectiveness while minimizing the costs. It can also be considered an optimization problem with

restrictions because it tries to optimize the QALYs without exceeding a budget constraint.

As we are working on a bicriteria problem, the analysis of each alternative will have two different outcomes: a cost in economic units and a effectiveness in QALYs. Given a ratio between both outcomes (economic units per QALY) we obtain a measure that represents the amount of money that a decision maker needs to pay in order to obtain a QALY.

If we compare two interventions A and B such that B is more effective than A ($E_B > E_A$) at a lower cost ($C_B < C_A$) we can say that the intervention B dominates the intervention A. In other situations, we have an alternative with a higher effectiveness but also at a higher cost. It is then necessary to calculate the ICER, defined as the ratio between the increment in cost and the increment in effectiveness.

$$ICER\,(A, B) = \frac{(C_B - C_A)}{(E_B - E_A)} \tag{3.3}$$

If we have a third intervention D, we can say that A and B jointly dominate D if the following conditions are met:

$$U_A < U_D < U_B \tag{3.4}$$

$$C_A < C_D < C_B \tag{3.5}$$

$$ICER\,(D, B) < ICER\,(A, B) < ICER\,(A, D) \tag{3.6}$$

In this case D can never be the optimal intervention as we can see in Figure 3.1.

The graphic representation of costs and effectiveness of different interventions is known as *cost-effectiveness plane.* Each point of the Cartesian space represents an intervention. The segmented line joining those interventions that are not dominated, either individually or jointly, is called the *efficiency frontier*, *cost-effectiveness frontier,* or *Pareto frontier* (Steuer, 1986). In Figure 3.2 we can see a more detailed example in which G is jointly dominated by A and B, and F is dominated by E.

$ICER\,(A, B)$ represents the money necessary to invest in the intervention B with

Figure 3.1: An example of extended dominance.

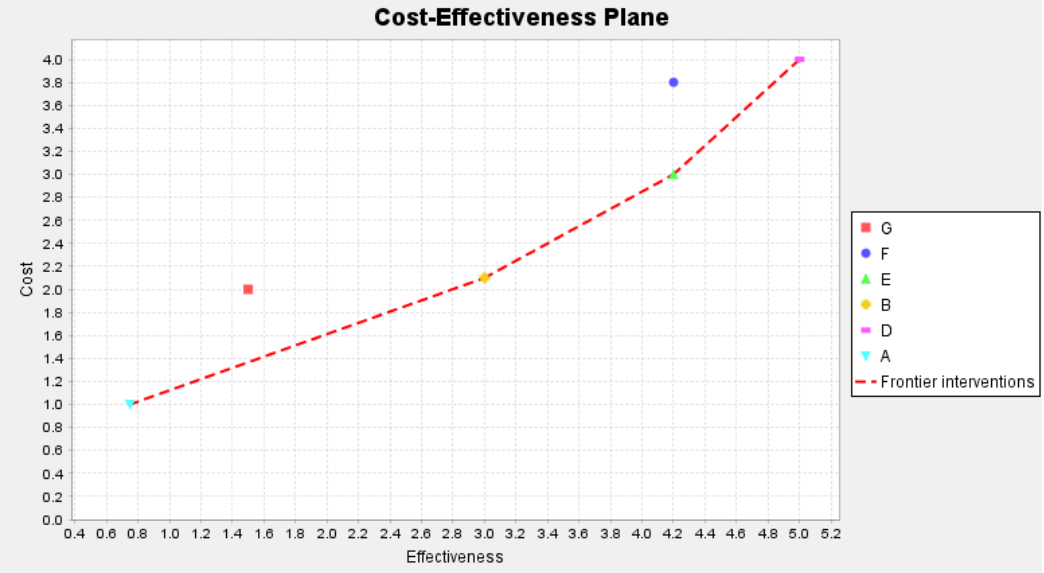


Figure 3.2: Cost-effectiveness frontier.

respect the intervention A to gain one more QALY. Given that WTP is the amount of money a decision maker is willing to pay in order to gain a QALY, we can say that B is cost-effective than A for a WTP (λ) when $ICER\,(A, B) < \lambda$.

3.1.4 Cost-effectiveness partitions

The result of the deterministic CEA can be presented as a Cost-Effectiveness Partition (CEP). A CEP contains, for every value of λ, a cost-effectiveness pair and the optimal intervention. This way the algorithm evaluates the tree for all the values of λ, grouping a finite set of intervals, such that within each interval the cost, the effectiveness, and the intervention are the same for all the values of λ. The thresholds that delimit the intervals are determined dynamically when evaluating the tree. This way the new roll-back algorithm can evaluate cost-effectiveness decision trees with embedded decision nodes. A *CEP* of n intervals (with $n \geq 1$) consists of the following elements:

- $\Theta = \{\theta_1, \ldots, \theta_{n-1}\}$ – a set of $n-1$ values (thresholds), such that $0 < \theta_1 < \ldots < \theta_{n-1}$,
- $C = \{c_0, \ldots, c_{n-1}\}$ – a set of n values (costs),
- $E = \{e_0, \ldots, e_{n-1}\}$ – a set of n effectiveness values, and
- $I = \{I_0, \ldots, I_{n-1}\}$ – a set of n interventions.

CEPs were introduced by Arias and Díez (2011) for the evaluation of DTs with embedded decision nodes, using the same algorithm as for unicriterion DTs (Algorithm 2.1), but assigning to each node a CEP instead of a single value. Similarly, the algorithms for the evaluation of unicriterion IDs can be adapted to the cost-effectiveness case (Arias and Díez, 2015). In Chapter 5, we will apply them to perform CEA with MIDs and DANs.

Figure 3.2 shows an example of CEA in which 4 interventions are not dominated (A, B, D and E) and 2 dominated (F and G). Each intervention is placed in the plane according to its cost and effectiveness (Table 3.1).

For each pair of consecutive interventions in the cost-effectiveness frontier, there is an ICER (a threshold in the CEP) which determines the slope of the line that

Intervention	Cost	Effectiveness
A	1.0	0.75
B	2.1	3.0
D	4	5.0
E	3.0	4.2
F	3.8	4.2
G	2.0	1.5

Table 3.1: Costs and effectiveness of the interventions of the example introduced in Figure 3.2.

λ	Cost	Effectiveness	Optimal intervention
0.0 - 0.49	1.0	0.75	A
0.49 - 0.75	2.1	3.0	B
0.75 - 1.25	2.1	3.0	E
1.25 - $+\infty$	3.0	4.2	D

Table 3.2: CEP intervals of the interventions of the example introduced in Figure 3.2.

connects the two interventions in the cost-effectiveness plane. In our example, the thresholds are:

$$\begin{aligned} ICER(A,B) &= 0.49 \\ ICER(B,E) &= 0.75 \\ ICER(E,D) &= 1.25\ . \end{aligned}$$

The resulting CEP is shown in Table 3.2.

3.2 Sensitivity analysis

Although we have a lot of useful information codified in our model all decisions are subject to some level of uncertainty. Each parameter of the model represents an

expected value that one variable takes in an specific scenario, but this values are usually extracted from non-perfect information, for example from studies that cannot analyze the entire population because it is unfeasible. When modeling a decision support system we must take into account all the uncertainty (Raiffa, 1968). The parameters determine the final output of the model and it is necessary to determine how its variation affect the outcomes. The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (Husereau et al., 2013) recommends Deterministic Sensitivity Analyses (DSAs) for individual parameters (e.g. discount rates) and Probabilistic Sensitivity Analyses (PSAs) to observe the impact of uncertainty across all the parameters. Health technology assessment guidelines require a PSA in all the studies that implies decision making (Weinstein et al., 2003; Claxton et al., 2005; Tan-Torres Edejer et al., 2012; NICE, 2013; Versteegh et al., 2016).

3.2.1 Representing the uncertainty

In the reference case, each parameter is an estimate value. Given the parameters of the model, each possible state is defined as a set of probabilities. Each of this possible states has associated an expected cost and an effectiveness. The expected costs and effectiveness of an intervention will be calculated as the multiplication of the probability of each possible state and the sum of the outputs expected for this state. So, the uncertainty of the parameters will propagate across the model and will affect the output of the reference case.

We can represent the uncertainty of a parameter in two ways: "forcing" a deterministic variation of its reference value or assigning a second order probability distribution that explains the uncertainty of the variable (Sucar, 2015). Depending on the type of the parameter, it is common to use a set of probability distributions to model the second order probability distribution (Briggs et al., 2006). Probability parameters could be fitted with beta distributions in order to ensure that the probability will still remain between 0 and 1. Costs are usually fitted with gamma or log-normal distributions and quality of life estimates with beta distributions.

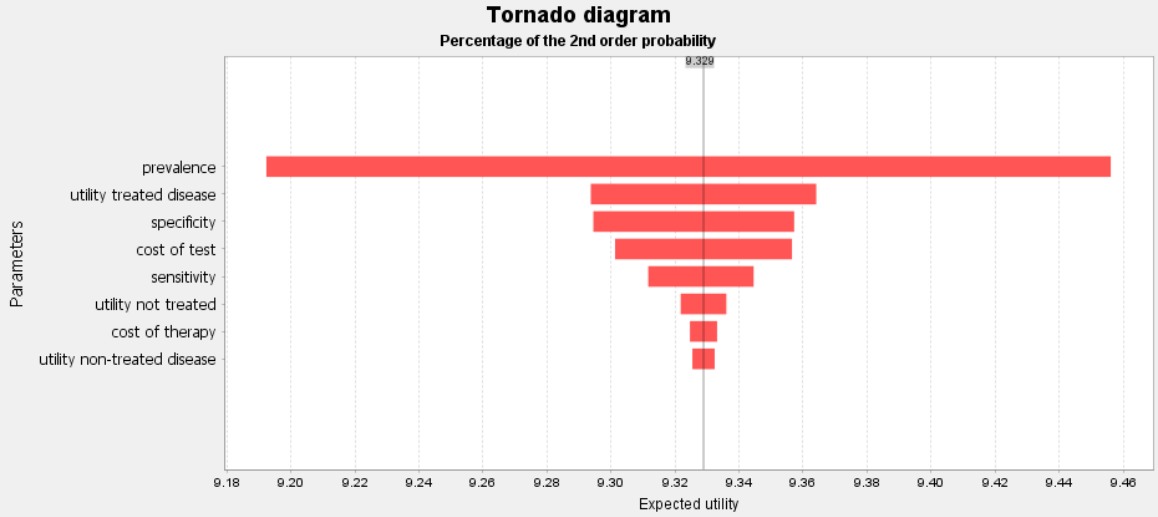


Figure 3.3: Example of a tornado diagram with 8 parameters.

3.2.2 Deterministic sensitivity analysis

DSA analyzes the sensitivity of the results to variations of one or more parameters. To carry a DSA the values of the parameters are manually modified across a specific range. The range of variation is commonly specified as a percentage variation of the reference value of this parameter. E.g. if we examine what happens to our results if the cost of a test vary on a fifth of its value ($\pm 20\%$) and the expected value of this test is €50, the range of variation goes from €40 to €60.

According the number of parameters that vary at the same time we can speak about univariate sensitivity analysis or multivariate sensitivity analysis. Univariate analysis varies only one parameter at each time assuming the principle of *ceteris paribus* (keeping all the other things parameters unchanged). Two of the most common plots used to present that information are tornado and spider diagrams.

Tornado diagram represent the impact of different parameters in the expected utility. For each parameter with uncertainty a univariate sensitivity analysis is performed. Tornado diagram is represented as a bar graph in which each bar corresponds to one parameter. The length of the bars is determined by the minimum and maximum expected utilities obtained when applying the variation to the parameters (see Fig. 3.3). The different bars are ordered following their impact to the expected utility, what gives the graph the look of a tornado.

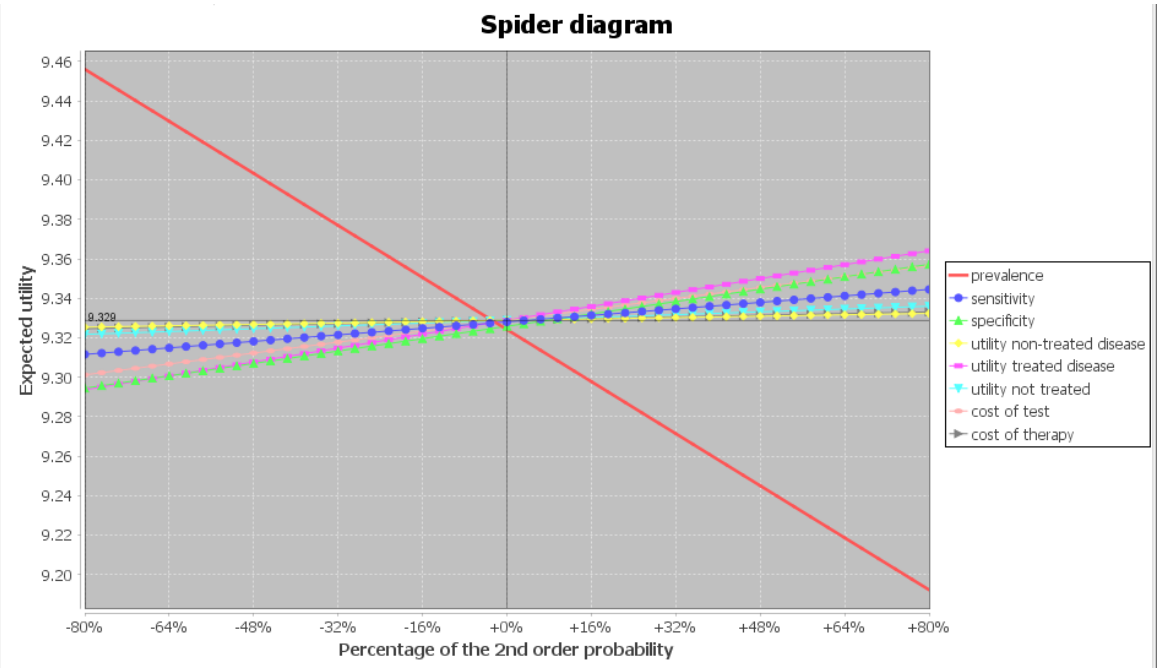


Figure 3.4: Example of a spider diagram with 8 parameters.

A limitation of tornado diagrams is that they only give information about the minimum and maximum expected utility but do not show whether the utility increases when the parameter increases or when it decreases, nor whether variations are linear. In contrast, spider diagrams (see Fig. 3.4) show how the expected utility varies with respect the variation of each parameter.

Although they are not so widely used, other sensitivity analyzes such the univariate utility plot graph (van der Gaag and Coupe, 2000) or the bivariate map graph (in which we can see the impact of the variation of two parameters at the same time) could be used to analyze and present useful information to the decision maker.

Another common sensitivity analysis tool in health economics is the Expected Value of Perfect Information (EVPI) which represents the WTP for obtaining perfect information, that is, how much we would be willing to pay for having perfect information. The EVPI is calculated as the difference between the expected value and the expected value given perfect information and could give us a notion about if it is worth to invest money to reduce the uncertainty of our model.

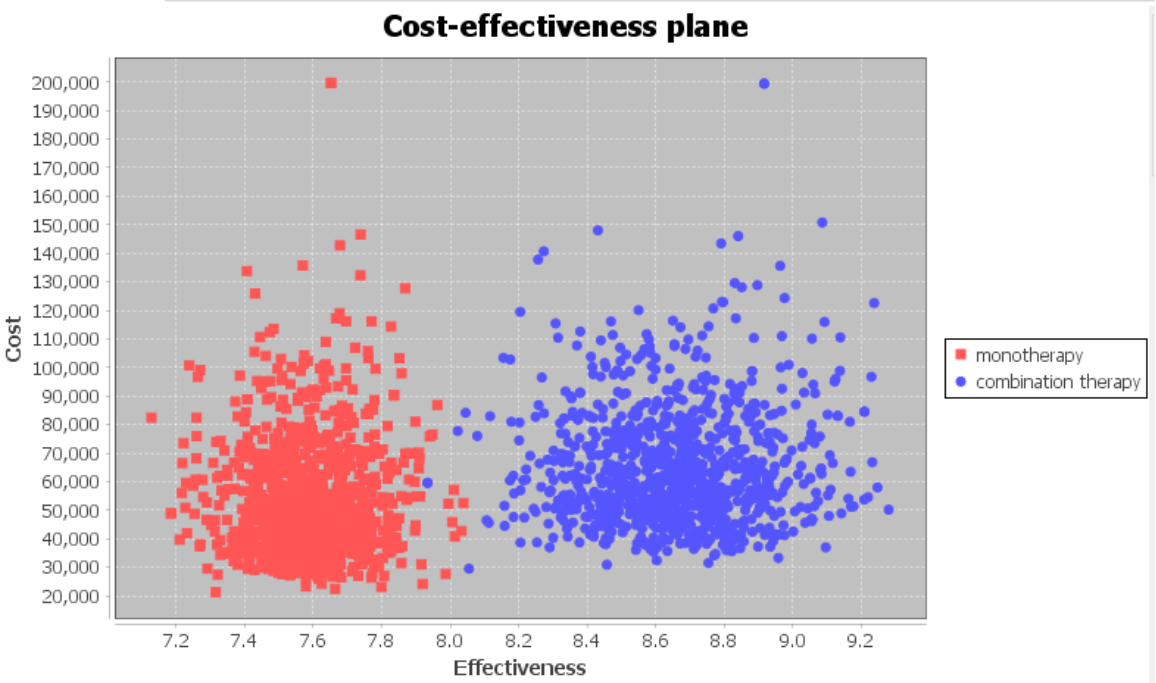


Figure 3.5: Example of a cost-effectiveness plane.

3.2.3 Probabilistic sensitivity analysis

PSA consists in analyzing the joint impact of all the uncertainties of the parameters of model. In this analysis we sample all the parameters according their second order probability distributions and perform an evaluation of the model obtaining the expected output. Using Monte Carlo simulation we run different sampled models and analyze the expected output of each simulation.

When applying PSA on a cost-effectiveness model, one of the most common visualizations is to represent each cost and effectiveness obtained from the Monte Carlo simulations in a cost-effectiveness plane. In Figure 3.5 we can see a graph in which each pair of points (red square and blue circle) represent one simulation. This representation express how concentrated or scattered are the simulations of the model, that giving us a clear view of the uncertainty of the model.

By analyzing the cost-effectiveness Monte Carlo simulations, we can calculate the percentage of simulations in which one intervention is cost-effective with respect the other for a range of specific WTP values. This information can be displayed on a

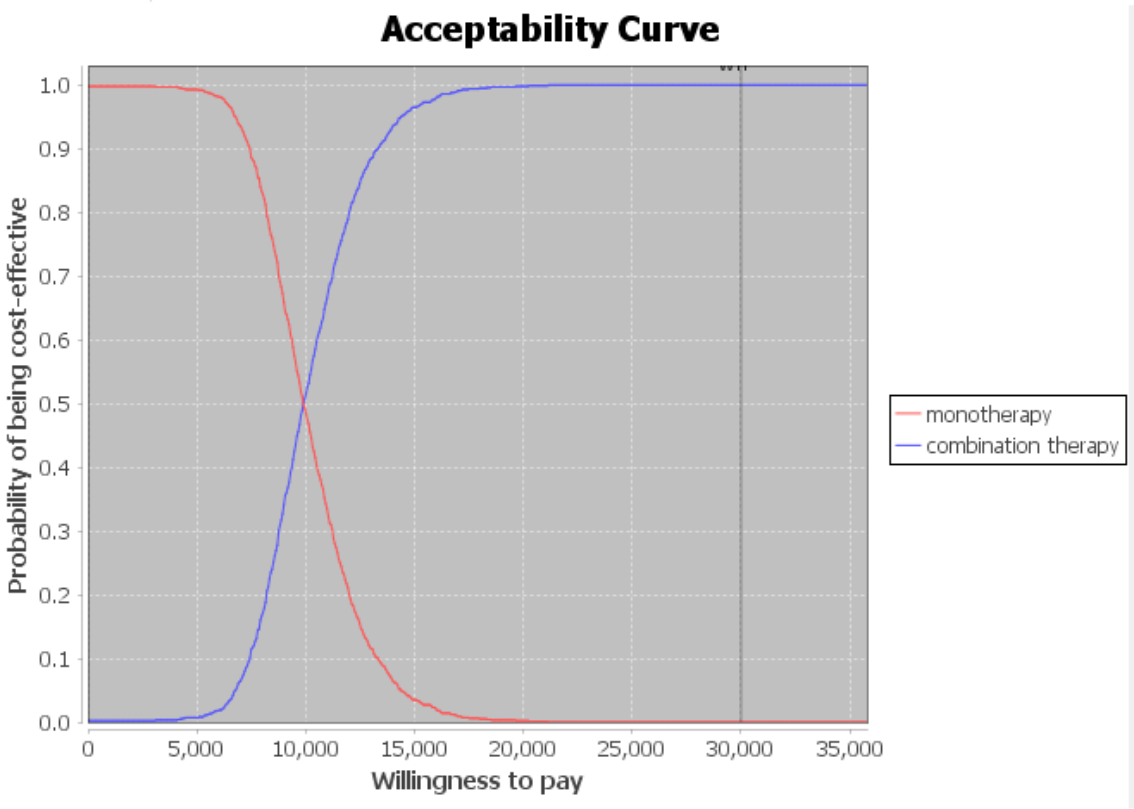


Figure 3.6: Example of an acceptability curve.

chart known as the cost-effectiveness acceptability curve and represents the probability of each intervention to be cost-effective than the other for each WTP. In the Figure 3.6 we can see that the cost-effectiveness acceptability curve that shows the probability of *combination therapy* being cost-effective with respect *monotherapy* for a range of WTP values.

Part III

Methodological contributions

4

Evaluation of Markov models with discontinuities

"All models are wrong, some are useful."

George E. P. Box

Results of this work were presented as a poster at the conference of the Society for Medical Decision Making (Pérez-Martín et al., 2017b) and later published in the journal of *Medical Decision Making* (Pérez-Martín et al., 2019).

4.1 Introduction

Markov models are the most popular modeling framework in health technology assessment (Beck and Pauker, 1983; Sonnenberg and Beck, 1993)—cf. Sec. 2.3.3. They represent the state of a patient as a set of mutually exclusive and exhaustive health states, such that the patient is in one and only one of them at a time. All the possible events are represented by transitions from one state to another. Every state has an associated cost and an effectiveness value. The total cost and effectiveness, accumulated over time, determine the net benefit of each intervention.

In economic evaluations of health technologies most Markov models use a discrete time approach, i.e., the time horizon is divided into a finite number of intervals of the same length, called *cycles*. In the classical presentation, transitions can only occur at the limit between consecutive intervals. The evaluation of a Markov model applies transition matrices to calculate for each cycle the probability that the patient is in one or another state. In the case of monotonically decreasing costs—which typically happens when patients are dying progressively—the assumption that transitions occur at the end of each cycle overestimates the cost, and assuming that they occur at the beginning underestimates it. The *half-cycle correction* (HCC) (Sonnenberg and Beck, 1993; Naimark et al., 2008) tries to minimize the error by adding a cycle of half duration at the beginning of the evaluation, which can be interpreted as assuming that transitions occur exactly in the middle of each cycle. Current guidelines for economic evaluation of health technologies recommend using this correction (Siebert et al., 2012). However, in recent years there has been a controversy about how to interpret and apply this method, and whether it should be replaced with a different approach, the *Life-Table (LT) method*, which averages the probabilities of state occupancy at the boundaries of each interval (Naimark et al., 2008; Naimark et al., 2013; Naimark et al., 2014; Barendregt, 2009; Barendregt, 2014). (Some authors present LTs as a way of implementing the HCC, but we refer to them as a different method for the sake of

clarity.)

More recently, Elbasha and Chhatwal (2016a) and Elbasha and Chhatwal (2016b) have proposed alternative *within-cycle corrections* based on numerical-integration techniques, which assume that cost and utility evolve continuously over time. These authors present an example whose results "suggest that the standard HCC method and the trapezoidal rule are not as accurate as Simpson's 1/3 and 3/8 rules" (Elbasha and Chhatwal, 2016a).

However, many medical models include abrupt changes in costs at certain points in time; for example, when expensive drug treatments are provided only for a limited time. The purpose of this work is to prove that different "corrections" applied to the same model may yield significantly different results. Using a slightly modified version of a model taken from the literature, we compare three "classical" approaches (those that assume that transitions occur at the beginning or at the end of each cycle, and the LT method) and several numerical-integration techniques. Our experiments show that when averaging the left and right limits of the cost at the point of discontinuity and applying the trapezoidal rule the result is more accurate than when applying the 3/8 Simpson rule to this model, and much more accurate than when applying numerical-integration techniques to the non-averaged model.

4.2 Empirical analysis

In order to compare several methods for the evaluation of Markov models, other authors have used synthetic examples (Soares and Canto e Castro, 2012; Naimark et al., 2013; Barendregt, 2014; Elbasha and Chhatwal, 2016a; Elbasha and Chhatwal, 2016b). In this work we use a slightly modified version of a real-world model, built by Chancellor et al. (1997) to determine the ICER of two interventions for HIV: monotherapy, which only applies zidovudine, and combination therapy, which adds lamivudine for two years, until it becomes ineffective for clinical reasons. The model has a cycle length of 1 year and was evaluated for a horizon of 20 years. It is now obsolete for clinical practice because there are more effective treatments for HIV, but it is still useful for pedagogic purposes. In particular, this model it is studied as an example in the book of Briggs et al. (2006). An Excel version of the original model is available at `www.herc.ox.ac.`

uk/downloads/decision-modelling-for-health-economic-evaluation. We have reimplemented it as an MID (Díez et al., 2017)—see Fig. 2.6—using OpenMarkov, an open source tool developed by our group (Arias et al., 2017a). The model is available at www.probmodelxml.org/networks.

The inaccuracy introduced by the discretization of time into cycles increases with the cycle length: when it approaches 0, all the methods studied in this work converge to the same results as a continuous-time model (Soares and Canto e Castro, 2012; Naimark et al., 2008; Naimark et al., 2013; Barendregt, 2014; Chhatwal et al., 2016; Elbasha and Chhatwal, 2016b). This should be our gold standard for comparing the accuracy of different methods. However, due to the difficulty of building a continuous-time version of Chancellor's model, we used as a gold standard a model with a cycle length of 1 day, which required a slight modification of the probabilities for the reasons explained in (Chhatwal et al., 2016). Figure 4.1 shows that the instantaneous cost function, calculated with this model, has a discontinuity at $t = 2$, when the patients in the combination therapy arm stop receiving lamivudine.

As mentioned in the introduction, we have applied three "classical" approaches. Two of them are based on the assumption that transitions occur at the beginning or at the end of each cycle. The third is the LT method (Barendregt, 2009; Barendregt, 2014). Another classical approach is the HCC, but it cannot be applied to this model because it assumes that the cost function for each state is constant.

We have also applied three of the numerical-integration methods proposed in (Elbasha and Chhatwal, 2016b): the trapezoidal rule and two Simpson rules, called 1/3 and 3/8. We also study the Riemann-sums rules but we do not include explicitly in the results because they are equivalent to the "classical" techniques that assume that transitions occur at the beginning or the end of each cycle. The classical approaches (transitions at the beginning of cycles, transitions at the end, and LTs) are insensitive to the value of the cost function at the point of discontinuity, $t = 2$, but numerical integration depends significantly on whether at this point we take the left or the right limit. We have examined both cases.

Finally, we have built a model in which the value of the cost *for each state* at $t = 2$ is manually set to the average between the values for $t < 2$ and for $t > 2$. This approach derives from the mathematical analysis in the section 4.3, which shows

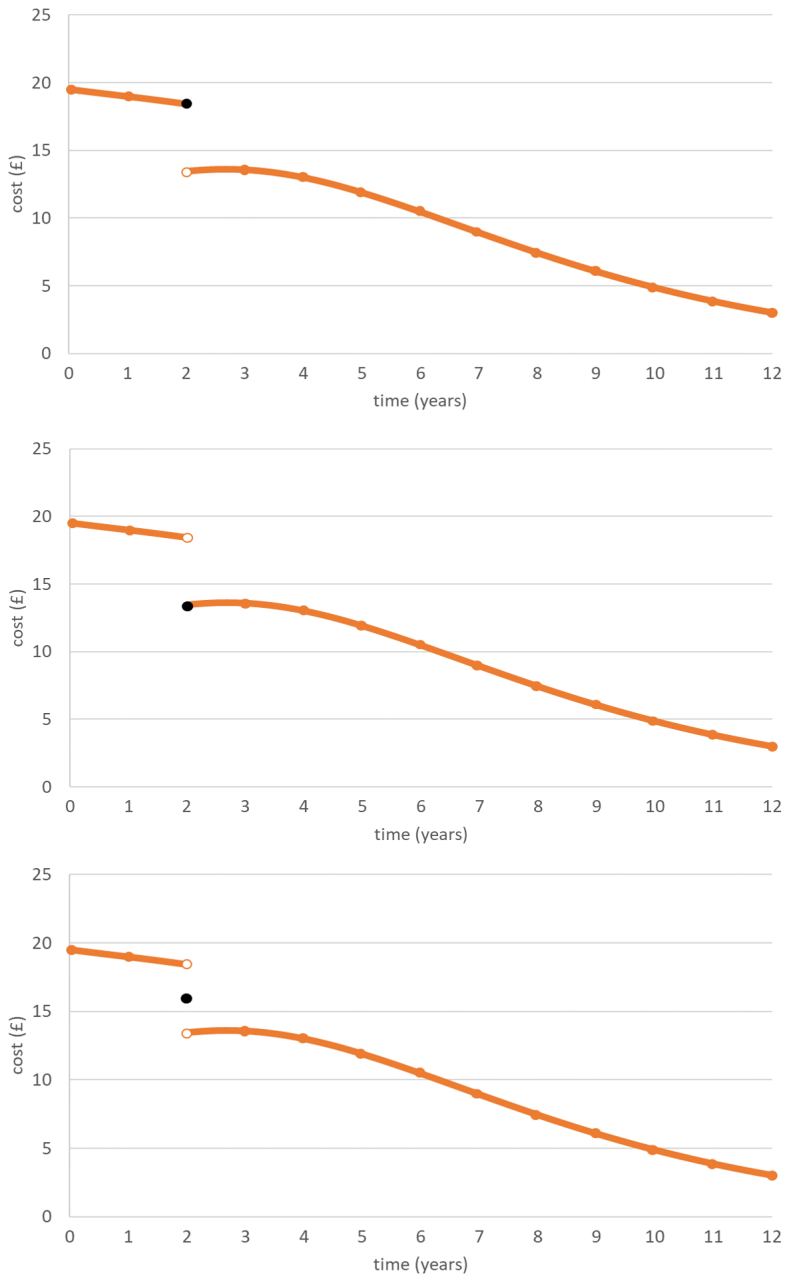


Figure 4.1: Instantaneous cost function for combination therapy obtained from our gold standard, i.e., the model with a cycle length of one day. It has a sharp discontinuity at $t = 2$. This figure shows the three possible scenarios: the point belongs to the first interval, the point belongs to the second interval, the point do not belong to any of these intervals.

that when the cost associated to some states has discontinuities at the boundaries between cycles but varies smoothly inside each cycle, the trapezoidal rule applied to the averaged model is a very good approximation. We have also applied other numerical-integration rules to the same model.

4.3 Mathematical analysis

A discrete Markov model consists of a set of states—usually a finite set—together with a probability distribution $P_0(s)$ for the states at time 0 and a function that describes the dynamics of the system and allows to compute the distribution $P_t(s)$ for any posterior moment. This is sometimes called *state membership function* because it indicates the probability that an individual is in state s at time t.

Markov models used for economic evaluation also contain functions that specify the cost and the effectiveness for every state at each moment, and the discounts for them. In this section we focus on how to estimate the total cost; the estimation of effectiveness is identical.

Let $c(s,t)$ be the cost function and $\gamma(t)$ the corresponding discount function; necessarily, $\gamma(0) = 1$. The instantaneous discounted cost is

$$c(t) = \sum_s P_t(s) \cdot c(s,t) \cdot \gamma(t) \tag{4.1}$$

and the total cost is

$$C = \int_{t_0}^{t_f} c(t) \cdot dt \,. \tag{4.2}$$

The instantaneous cost is measured in monetary units per time unit (for example, in euros per year) and the total cost in monetary units. In most models the cost function $c(t)$ decreases monotonically because the cohort that enters the model requires fewer and fewer resources as the patients progressively die.

Discrete-time Markov models divide the time into a finite number of intervals of the same length, τ, called *cycles*. When the model is evaluated for a limited number of cycles, h (the horizon), the time points that delimit the intervals are $\{0, \tau, 2\tau, \ldots, h\tau\}$.

The probability of being in state s at time $i\tau$ is

$$P_i(s) = \sum_{s'} P_i(s \mid s')\, P_{i-1}(s') \ . \tag{4.3}$$

where $P_i(s \mid s')$ is the *transition matrix*. When the transition matrix is time-independent, we can drop the subindex and just write $P(s \mid s')$. (In order to simplify the notation, we have written $P_i(s)$ instead of $P_{i\tau}(s)$.) This way we obtain a set of probability distribution, $\{P_0(s), P_1(s) \ldots, P_h(s)\}$.

4.3.1 Classical approaches

The classical approach to the evaluation of discrete-time Markov models assumes that transitions between states can only occur at the limit between cycles, i.e., the state of the system does not change within each cycle. It also assumes that the cost function for each state is constant in the interval $(i\tau, (i+1)\tau)$; let $c_i(s)$ be this cost. Additionally, the discount function usually decreases very slowly, which means its value in this interval its very approximately $\gamma(i\tau)$. The accuracy of this approximation depends on the cycle length.

Some models assume that transitions only occur at the end of a cycle, i.e., the probability in the interval $[i\tau, (i+1)\tau)$ is $P_i(s)$. With these assumptions and approximations, the total cost is

$$C_E = \sum_{i=0}^{h-1} \sum_{s} P_i(s) \cdot c_i(s) \cdot \gamma(i\tau) \cdot \tau \ , \tag{4.4}$$

where the subindex E stands for "end". When the cost function (cf. Eq. 4.1) decreases monotonically, this assumption leads to an overestimation of the total cost, i.e.,C_E is higher than the true cost, C, given by Equation 4.2.

Alternatively, it is possible to assume that transitions only occur at the beginning of each cycle, i.e., the probability in the interval $(i\tau, (i+1)\tau]$ is $P_{i+1}(s)$. Therefore,

$$C_B = \sum_{i=0}^{h-1} \sum_{s} P_{i+1}(s) \cdot c_i(s) \cdot \gamma(i\tau) \cdot \tau \ , \tag{4.5}$$

where the subindex B stands for "beginning". When the cost function decreases monotonically, this assumption leads to an underestimation of the total cost, i.e., $C_B < C < C_E$.

4.3.1.1 The trapezoidal rule and the half-cycle correction

In an attempt to obtain a better approximation, the trapezoidal rule computes the average cost for each cycle as the arithmetic mean of the instantaneous costs at its boundaries:

$$C_{TR} = \sum_{i=0}^{h-1} \frac{c(i\tau) + c((i+1)\tau)}{2} \cdot \tau \tag{4.6}$$

$$= c(0) \cdot \frac{\tau}{2} + \sum_{i=1}^{h-1} c(i\tau) \cdot \tau + c(h\tau) \cdot \frac{\tau}{2} \,. \tag{4.7}$$

This equation was applied when evaluating Markov DTs with different software packages, such as SMLTREE, Decision Maker, and TreeAge (Sonnenberg and Beck, 1993; Naimark et al., 2013). Unfortunately, when explaining this method those authors did not present it as the trapezoidal rule applied to instantaneous costs, but as a way of approximating the state-occupancy probabilities in order to subsequently estimate the cost accrued in each cycle. Furthermore, sticking to the traditional assumption that transitions can only occur at certain moments, they interpreted the first term in Equation 4.7 as the cost accrued during a cycle of length $\tau/2$ in which no transition occurs; hence the name "half-cycle correction" (HCC) and the assertion that this "correction" is equivalent to assuming that the transitions occur at the time points $\{0.5\tau, 1.5\tau, 2.5\tau, \ldots\}$, i.e., halfway through each cycle (Beck and Pauker, 1983; Naimark et al., 2008).

In summary, the trapezoidal-rule method and HCC apply exactly the same equation but with a different interpretation.

4.3.1.2 The life-table method

An alternative approach, based on the same premises as the usual presentation of the HCC, consists in first averaging the state-occupancy probabilities at the boundaries

and then calculating the costs:

$$C_{LT} = \sum_{i=0}^{h-1} \sum_{s} \frac{P_i(s) + P_{i+1}(s)}{2} \cdot c_i(s) \cdot \gamma(i\tau) \cdot \tau \,, \tag{4.8}$$

where $c_i(s)$ is the cost inside the i-th interval, again assumed to be constant. This is called the LT because it is based on the procedure that demographers use it to estimate life expectancy (Barendregt, 2009).

It is easy to check that

$$C_{LT} = \frac{C_E + C_B}{2} \,.$$

We have already seen that when the total-cost function decreases monotonically, C_E is an overestimation of the true cost and C_B is an underestimation, so their arithmetic mean, C_{LT}, is expected to be closer to the true value than if we assumed that all transitions occur either at the beginning or at the end of each cycle.

4.3.1.3 A comparison of HCC and LT

We have seen that HCC was explained as a method that approximates the state-occupancy probability within each interval by applying the trapezoidal rule to the state-occupancy probability at the boundaries between cycles, but it was implemented by applying Equation 4.7 to the cost, assumed to be constant, accrued in each cycle. This inconsistency motivated the severe criticism of (Barendregt, 2009; Barendregt, 2014).

However, we should first note that in general the two methods yield almost the same results, despite the apparent differences in the way of doing the calculations. In fact, when the cost function is time-independent for each state (a basic assumption of the HCC), we can write $c(s)$ instead of $c(s,t)$ in Equation 4.1, so that Equation 4.6 leads to

$$C_{HCC} = \sum_{i=0}^{h-1} \sum_{s} \frac{P_i(s) \cdot c(s) \cdot \gamma(i\tau) + P_{i+1}(s) \cdot c(s) \cdot \gamma((i+1)\tau)}{2} \cdot \tau \,. \tag{4.9}$$

Comparing this with Equation 4.8, we observe that they only differ in the way of applying the discounts. If there is no discount, $C_{LT} = C_{HCC}$. If the cycle length

τ is short or the discount function decreases slowly, then $\gamma(i\tau) \approx \gamma((i+1)\tau)$ and $C_{LT} \approx C_{HCC}$, but HCC is more accurate because it uses the values of $\gamma(t)$ at the boundaries of each cycle, in coherence with the trapezoidal rule, instead of applying the right-boundary discount all along the cycle.

Additionally, the HCC approach presented here only requires that the total cost function, $c(t)$, varies smoothly inside each cycle, while LT requires that the cost of each state, $c(s,t)$, is constant inside each cycle. When this condition does not hold, HCC is more accurate than LT.

For these reasons, we disagree with *some* of the arguments claiming the superiority of LT over HCC. For example, Barendregt (2009) argued the standard HCC method is incompatible with discounting; Naimark et al. (2013), who were initially strong advocates of the HCC, finally agreed with him. However, the instantaneous cost in Equations 4.6 and 4.7 clearly include the discounts (see Eq. 4.1).

Barendregt (2009) also said: "I know of very few relevant Markov models in medical decision making where QALY weights and unit costs are constant across all cycles". However, HCC does not require that the cost function $c(t)$, given by Equation 4.1, be constant. In our analysis, the derivation of Equation 4.7 only required that $c(s,t)$ be continuous (for every state s).

In turn, Naimark et al. (2013) said that "the standard approach to the HCC assumes that the state membership curve is declining and monotonic." Our derivation of Equation 4.7 does not require that assumption, so we disagree with their assertion that "the standard HCC is not appropriate for nonmonotonic states", which they define as the states whose membership function (their state-occupancy probability) is not monotonic.

In summary, we claim that there was nothing wrong in the application of the HCC—in fact, Equation 4.7 is in general more accurate than 4.8. The problem was in the way of explaining and justifying the method.

4.3.2 Numerical-integration approaches

4.3.2.1 The approach of Naimark et al.

To our knowledge, Naimark et al. (2013) were the first to propose using numerical-integration techniques to obtain an approximation similar to that of the HCC but more accurate. Using our mathematical notation, their approach is the result of combining Equations 4.1 and 4.2 and permuting the summation and the integration, so that when the cost function for each state is constant, the total cost is

$$C = \sum_s c(s) \cdot P_{\text{accrued}}(s) \ , \tag{4.10}$$

where

$$P_{\text{accrued}}(s) = \int_{t_0}^{t_f} P_t(s) \cdot \gamma(t) \cdot dt \ . \tag{4.11}$$

They proposed computing this integral (for each state s) using Simpson's simple rule.

Clearly, the assumption that $c(s,t)$ is time-independent makes this method unsuitable for discontinuous cost functions.

4.3.2.2 The approach of Elbasha and Chhatwal

Elbasha and Chhatwal (2016a) and Elbasha and Chhatwal (2016b) also used numerical integration, but in a different way: they depart from the set of probability distributions for the limits between intervals, $\{P_0(s), \ldots, P_h(s)\}$, as in the classical approach, but the probabilities inside each interval are not assumed to be constant; they are just unknown. Instead of trying to estimate the accrued probability for each state, as in the approach above, Elbasha and Chhatwal compute the instantaneous cost at the boundaries between intervals, $\{c(0), c(\tau), c(2\tau), \ldots, c(h\tau)\}$, and then apply different numerical-integration techniques. They prove that the left Riemann sum is equivalent to assuming that the transitions occur at the end of each cycle, the right Riemann sum is equivalent to assuming that they occur at the end, and the trapezoidal rule is equivalent to the HCC.[1]

[1] Elbasha and Chhatwal (2016b) also say that the trapezoidal rule will give the same result as LT. However, we have already explained why HCC and LT give different results when there are

4.3.3 The trapezoidal rule for models with discontinuities

Both HCC and numerical-integration methods implicitly assume that the cost function is continuous. When applying the trapezoidal rule, which is equivalent to the HCC, the cost accrued in the i-th cycle is

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \frac{c(i\tau) + c((i+1)\tau)}{2} \cdot \tau \,. \tag{4.12}$$

If the cost function varies smoothly inside the interval but there are discontinuities at the boundaries, we should write instead

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \left(\lim_{t \to i\tau^+} c(t) + \lim_{t \to (i+1)\tau^-} c(t) \right) \cdot \tau \tag{4.13}$$

and Equation 4.7 should be replaced with

$$C_{TR} = c(0) \cdot \frac{\tau}{2} + \sum_{i=1}^{h-1} c^*(i\tau) \cdot \tau + c(h\tau) \cdot \frac{\tau}{2} \,, \tag{4.14}$$

where

$$c^*(i\tau) = \sum_s P_i(s) \cdot \frac{1}{2} \left(\lim_{t \to i\tau^-} c(s,t) + \lim_{t \to i\tau^+} c(s,t) \right) \cdot \gamma(i\tau) \,. \tag{4.15}$$

This means that if the cost function $c(s,t)$ is discontinuous at $i\tau$ for a state s, we can still apply the trapezoidal rule but in the calculation of the "instantaneous cost" at $i\tau$ we must take the average of the left and right limits for $c(s,t)$. The adjustment can be implemented by setting $c(s, i\tau)$ to this average in the model.

This analysis justifies the application of the trapezoidal rule to the averaged model, which we recommend as the most accurate approach for evaluating Markov models with discontinuities.

discounts. Section 9.2.1 in (Gray et al., 2011) shows an example in which the numerical results yielded by both methods are different.

Method	ICER (£/QALY)	Percentage error (%)
Gold standard (cycle length = 1 day)	6,543	
Classical approaches		
transitions at the beginning of cycle	6,680	2.09
transition at end of cycle	6,401	−2.17
life tables	6,539	0.06
Numerical integration		
– **Left limit at $t = 2$ years**		
trapezoidal	7,440	13.71
1/3 Simpson	7,146	9.22
3/8 Simpson	7,603	16.20
– **Right limit at $t = 2$ years**		
trapezoidal	5,640	−13.8
1/3 Simpson	5,945	−9.14
3/8 Simpson	5,579	−14.73
– **Average of limits at $t = 2$ years**		
trapezoidal	6,540	−0.05
1/3 Simpson	6,545	0.03
3/8 Simpson	6,590	0.72

Table 4.1: Impact of different within-cycle correction methods in the ICER and the percentage error with respect to gold standard.

4.4 Results of the empirical analysis

Table 4.1 summarizes the results obtained with each method. As expected, the approaches that assume that transitions occur at the beginning or the end of each cycle give higher errors, but it would have been difficult to predict their signs because biases affect the cost and the effectiveness of each of the interventions. Among the three classical approaches, the LT method yielded the smallest error, as expected. In contrast, the error introduced by numeric-integration techniques may be higher than 13% with respect to the value computed with the gold-standard model.

We also observe in the table that numerical-integration methods are very sensitive to the value of the cost functions at the point of discontinuity. It was expected, because these methods try to estimate the value of the cost, $c(t)$, around the point $t = 2$ by "propagating" it towards its left and its right, thus amplifying the effect of the choice

made for $c(2)$. The best results are obtained when taking the average of the left and right limits for $t = 2$, but these results are not much better than when using the LT method. It is also interesting to note that, even though the 3/8 Simpson rule is usually more accurate than the trapezoidal rule and the 1/3 Simpson rule in the absence of discontinuities, in this example it is clearly the opposite.

With respect to computational efficiency, the gold standard model, with daily transitions, was evaluated in 22.64 hours, while the classical approaches and the numerical-integration methods applied to a model with yearly transitions only took one or two seconds (1.40 s on average).

4.5 Discussion

The HCC was proposed by Sonnenberg and Beck (1993) as a method for computing accumulative outcomes (for example, cost and effectiveness) more accurately in discrete-time Markov models. Naimark et al. (2008) offered two analytical justifications of the HCC with didactic purposes, but their idea was criticized as a "kludge" in a paper entitled "The half-cycle correction: banish rather than explain it" (Barendregt, 2009), which argued that the HCC should be replaced with a more accurate technique, the *LT method*. A few years later Naimark et al. (2013) proposed several modifications aimed at "redeeming the kludge", but again Barendregt (2014) criticized their work severely, to the point that Naimark et al. (2014) surrendered and accepted that "the standard approach to the HCC is flawed and should be abandoned".

In our opinion, the problem was that in (Sonnenberg and Beck, 1993) and (Naimark et al., 2008) the HCC was not justified as the application of the trapezoidal rule to the instantaneous discounted cost but as its application to the state-occupancy probabilities in order to subsequently calculate the cost and effectiveness accrued in each cycle. That was the source of several mathematical inconsistencies and made the HCC impossible to apply when the cost function is discontinuous (Barendregt, 2009; Barendregt, 2014). Another problem of the traditional way of presenting the HCC is the assumption that transitions can only occur at the boundary between cycles—an idea exposed in most papers and virtually all the textbooks that explain Markov models. However, following the ideas of Elbasha and Chhatwal (2016a) and

Elbasha and Chhatwal (2016b), we have shown that (in the absence of discontinuities) the trapezoidal rule is formally equivalent to the HCC—see Sec. 4.3.1.1—but the interpretation is different, because it does not assume that transitions only occur at the boundaries between cycles or halfway through each cycle.

Elbasha and Chhatwal also argued that more sophisticated numerical-integration techniques, such as Simpson 1/3 and 3/8 rules, in general give more accurate results than the trapezoidal when the function of interest is not linear inside each cycle. Unfortunately they did not take into account one of Barendregt's criticism of the HCC: its inaccuracy when the model has discontinuities, a problem which also affects numerical-integration approaches. We faced it when evaluating Chancellor's model: the cost function has a severe discontinuity at the end of the second year, when lamivudine becomes clinically ineffective and is withdrawn.[2] Common sense says that it does not matter whether it is withdrawn a second before $t = 2$ or one second later, so a model in which lamivudine is applied in the interval $[0, 2)$ should give the same results as if it is applied in the interval $[0, 2]$. A continuous-time model would be insensitive to this modeling decision, but the evaluation of a discrete-time model with a large cycle length is very sensitive to the cost at $t = 2$ because this is an input of the numerical-integration algorithm. In our example, the difference between withdrawing lamivudine just before or after $t = 2$ is higher than 27%; the error with respect to the gold standard is $\pm 13\%$. Interestingly, the 3/8 Simpson rule, which is generally the most accurate for continuous functions, gave worse results than the trapezoidal rule and the 1/3 Simpson rule; the results were even worse than when assuming that all transitions occur at the beginning or the end of a cycle. If the ICER estimated is close to the WTP threshold, this error may lead to making a wrong decision.

As mentioned above, the errors were reduced when applying numerical-integration techniques on a model in which the cost at $t = 2$ is the average between administering lamivudine and not administering it. The application of the trapezoidal rule to this model is justified by the algebraic analysis in the sec. 4.3.3 and results in an error of only -0.05% in the ICER. The 1/3 Simpson rule also gives a small error, 0.03%. These are smaller than the 0.06% of the LT method and much smaller than the 0.72% of the 3/8 Simpson rule.

[2]The cost function $c(s, t)$ is discontinuous only for the three states in which the patient is alive. When the patient is dead (fourth state) there is no discontinuity.

A limitation of our study is that we have only studied one model. Further studies are necessary in order to determine to what extent the qualitative results obtained generalize to other models. However, our analysis serves at least as a warning that in the case of discontinuities numerical-integration techniques may give wrong results because they were designed for continuous functions.

4.6 Conclusion

In many cases it is not possible (or its unfeasible given the computation times) to build a continuous-time Markov model or to shorten the cycle length of a given discrete-time model (Chhatwal et al., 2016), but within-cycle corrections, such as the HCC, may give very good approximations (Soares and Canto e Castro, 2012; Elbasha and Chhatwal, 2016a; Elbasha and Chhatwal, 2016b). However some models have discontinuities in costs due to the withdrawal of expensive therapies. (There might also be discontinuities in the effectiveness, but we have not found any example.) The LT method is insensitive to discontinuities at the boundaries between cycles, but the HCC cannot be applied because it is based on the assumption that the cost for each state is constant. Numerical-integration techniques may also lead to significant errors in this case. We have proved mathematically that the trapezoidal rule—formally equivalent to the HCC but with a different interpretation—yields a good approximation also when the cost function has discontinuities at the boundaries between cycles but varies smoothly within each cycle, provided that it is applied on a model in which the value of the instantaneous cost function for each point of discontinuity is set to the average of the left and right limits. In the real-world model we have studied, the 1/3 Simpson rule also gave a good result, even though we did not have mathematical justification for it, but the 3/8 Simpson rule, which is more accurate for models without discontinuities, introduced significantly larger errors. As a conclusion, when a model has discontinuities we recommend building an averaged model and applying the trapezoidal rule instead of more sophisticated numerical-integration techniques.

5

Cost-effectiveness evaluation with Probabilistic Graphical Models

"A little inaccuracy sometimes saves a ton of explanation."

Saki

This chapter presents the methodological contributions to the cost-effectiveness evaluation with two types of PGMs: MIDs and DANs. Results of the work carried out in this chapter has been included in Díez et al. (2017), Arias et al. (2017a), Arias et al. (2017b), and Arias et al. (2019).

5.1 Introduction

Several types of computational models are available for medical decision making. As we mentioned in Section 2.3.1, DTs (Raiffa, 1968) are the most popular, but they are suitable only for small problems because their size grows exponentially with the number of decisions and chance variables. When doing CEA, the standard roll-back algorithm cannot evaluate DTs with embedded decision nodes—i.e., those that are not the root of the tree (Arias and Díez, 2014). A possible solution consists in having only one decision node (at the root of the tree) representing all the possible strategies, as recommended by Kuntz and Weinstein (2001), but this solutions usually leads to huge DTs even for small problems (Arias and Díez, 2011). A much more efficient algorithm proposed by Arias and Díez (2011) can perform CEA on trees with embedded decision nodes, but still the size of problems that can be represented is very limited.

Standard algorithms for IDs, such as arc reversal (Olmsted, 1983; Shachter, 1986) and variable elimination (Jensen and Nielsen, 2007), can only evaluate unicriterion IDs, which makes them unsuitable for health technology assessment. Arias and Díez (2015) and algorithm for perform CEA with IDs, which is an adapted version of their algorithm for DTs, and applied it to two models whose equivalent DTs would contain more than 10,000 leaves.

CEA is even more difficult when time must be modeled explicitly. The most common approach is to build state-transition models (Briggs et al., 2006; Gray et al., 2011; Siebert et al., 2012), which allocate members of a population into one of several categories, or health states, and discretize time into a set of fixed-length intervals, called cycles. Markov DTs, originally called Markov cycle trees (Sonnenberg and Beck, 1993), were the first representation used to implement state-transition models. There are several software used for building state-transitions models (such as Excel or TreeAge), but it is also common to find models written in programming languages

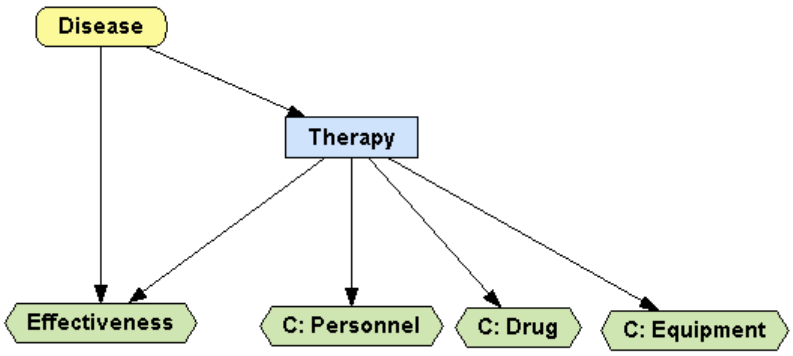


Figure 5.1: Example of multicriteria MID.

such as C++ or R (Tosh and Wailoo, 2008).

In this chapter we present the algorithms developed for CEA with two types of PGMs proposed by our group: MIDs, for temporal problems, and DANs, for asymmetric models, in particular models with partially ordered decisions.

5.2 Cost-effectiveness analysis with Markov influence diagrams

When I started this thesis, the researchers at the CISIAD had already proposed a new type of PGM, called Markov Processes with Atemporal Decisions (MPAD), which was designed to perform CEA with state-transition models. This section presents the modifications on which I worked and that were necessary to improve MPADs, including the one that contributed to changing its name to Markov Influence Diagrams (MIDs). Please note that in this section we will use the acronym MPAD for the formalism that already existed and MID for the formalism to which I contributed.

5.2.1 Multicriteria analysis

MPADs allowed to define two criteria: cost and effectiveness. To perform the analysis, each value variable of the model had to be assigned to one of these criteria. At that time, MPADs only supported two criteria, cost and effectiveness.

MIDs support multiple criteria, each one measured in its own units. In Figure 5.1

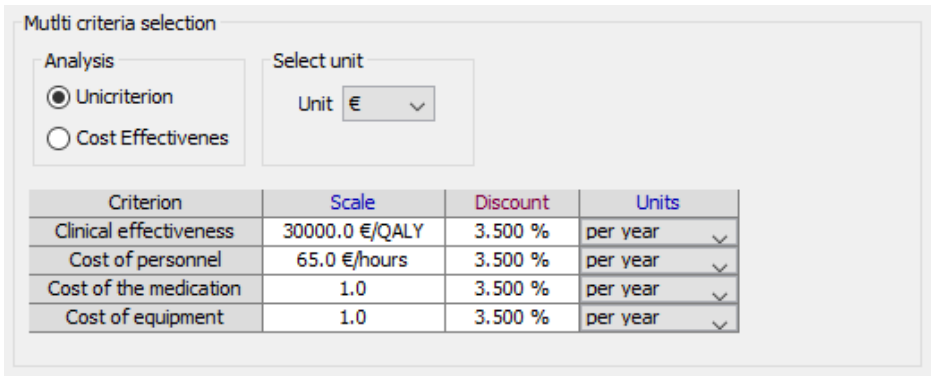


Figure 5.2: Unicriterion evaluation in a multicriteria model.

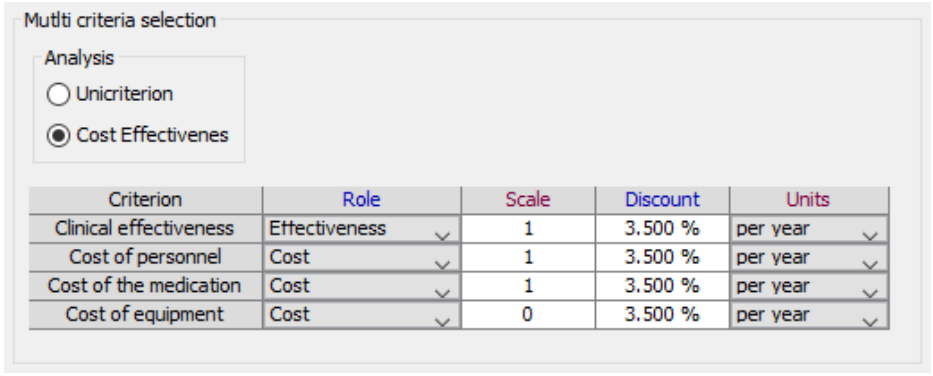


Figure 5.3: Cost-effectiveness evaluation in a multicriteria model.

we can see a model with four different criteria: clinical effectiveness, cost of personnel, cost of the medication, and cost of equipment. Each criteria has its own measurement units; for example, we can quantify the cost of the personnel in hours of work, and the cost of the drug in economic units.

There are at least two ways to evaluate MIDs with more than two criteria. First, we can combine all the criteria into a single one using a weight for each criterion, as shown in Figure 5.2, and then perform an unicriterion analysis. Alternatively, we can reduce the original criteria to only two criteria, cost and effectiveness, as shown in Figure 5.3, and then perform a CEA.

To perform a CEA, all the decision criteria must have an assigned role, either cost or effectiveness, depending on whether they should be minimized or maximized. In this particular type of bicriteria analysis, the factor of each criterion is used to obtain weighted values. This is useful when we want to have multiple scenarios in a

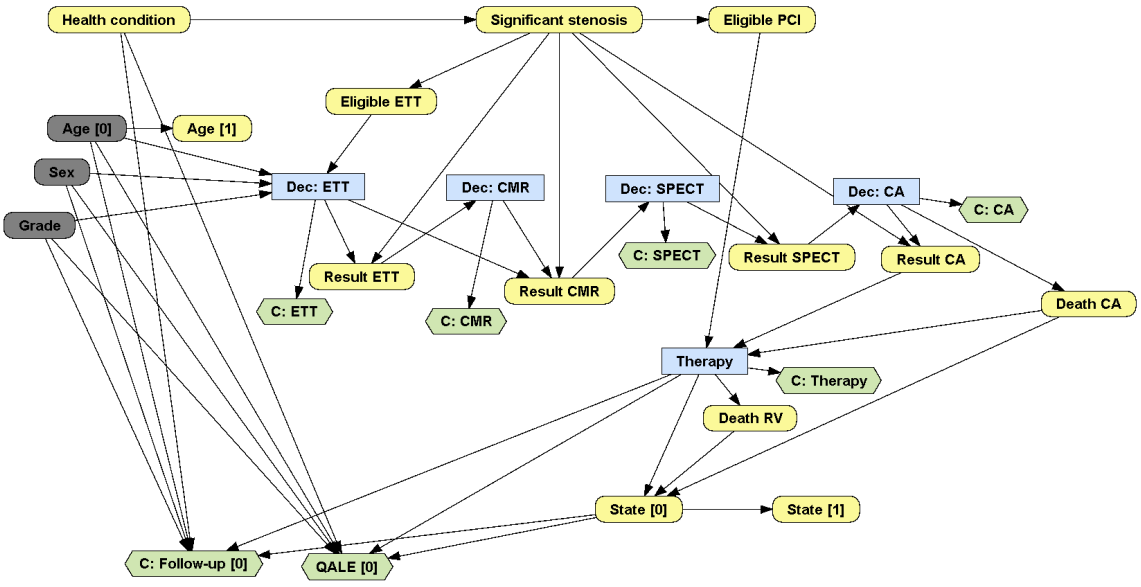


Figure 5.4: Example of mulidecision MID.

single model, for example, if we want to not include some expected outcomes we can assign factor equals to zero to its criterion. In Chapter 6 we will use these decision criteria weights to perform two analyses with the same model: one including social costs, and one without including them.

5.2.2 From MPADs to MIDs

MPADs could contain multiple atemporal decisions, but they had to be merged into a single one with the Cartesian product of their states. This approach, is similar to the one proposed by Kuntz and Weinstein (2001) for DTs. The main limitation of this evaluation is that it does not allow to have evidence, neither before the decisions, nor between them. In contrast, Figure 5.4 shows a MID in which there is evidence before the decisions (e.g. *Eligible ETT*), which can be used to particularize the CEA to specific subpopulations, and evidence between them (e.g. *Result ETT*), which may represent the results of different test.

Even though the CISIAD had developed specific algorithms for MPADs, we realized that it was possible to convert them into IDs by applying several transformations, such as expanding the network or weighting each criterion. Therefore, these new type

of model can be understood as an extension of IDs an be evaluated with a modified version of the cost-effectiveness algorithms for IDs, which allow the models to represent more complex problems (for example the one presented in Figure 5.4). The integration of MPADs and IDs results in a new type of model that we decided to call MIDs. This integration required several changes in our open-source software, OpenMarkov, which are described in Appendix C.

To evaluate the resulting ID, value nodes should be kept separate until the moment of eliminating the first decision, because if we joined them into a single node when initializing the algorithm, the computational complexity would grow exponentially with the horizon.

5.2.3 Efficient evaluation of MIDs

The evaluation of a MID can take a long time and consume a lot of memory due to the size of the equivalent ID, which is proportional to the temporal horizon. A good elimination heuristic can reduce the time and memory spent in the evaluation but it is not always enough to evaluate large complex models.

In Chapter 4 we explained the need of calculating the cost and effectiveness in each cycle of a Markov model. With the algorithms we had implemented for MIDs we could perform CEA, obtaining the CEPs; however in order to obtain the cost and effectiveness for each cycle, it was necessary to evaluate the network once for each value-variable in the expanded model; for example, the model in Figure 5.4, which has a temporal horizon of 20 years, must be evaluated 45 times: 20 times for each of the two temporal value-variables and once for each atemporal value-variable. Given that this approach consumes a lot of time and memory, we developed Algorithm 5.1, which adds and deletes nodes dynamically taking advantage of the Markov property (see Sec. 2.2.3), while storing the cost and effectiveness of each cycle.

As we can see in Figure 5.5, each iteration takes into consideration all the variables relevant at this moment. After building the network for a particular time slice, the algorithm removes the variables that are not relevant for future slices, thus obtaining an intermediate network that only contains the parents of the sub-networks in the next slices. This iterative evaluation returns the cost and effectiveness of each

Algorithm 5.1: Temporal evaluation of MIDs

Input: A Markov influence diagram (*MID*)

Result: Set of expected instantaneous costs and effectiveness in each cycle, c_i and e_i

```
network ← getAtemporalVariables(MID)
for slice ← 0 to temporalHorizon do
    network.addVariablesInSlice(MID, slice)
    network.removeUnnecesaryVariables()
    c_i, e_i ← network.removeVariables()
```

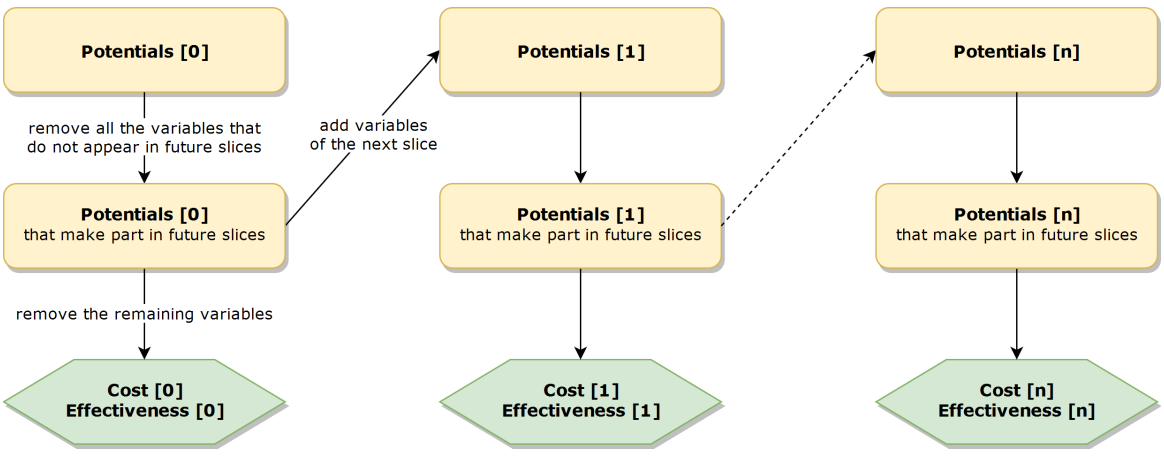


Figure 5.5: Flowchart of the temporal-evaluation algorithm.

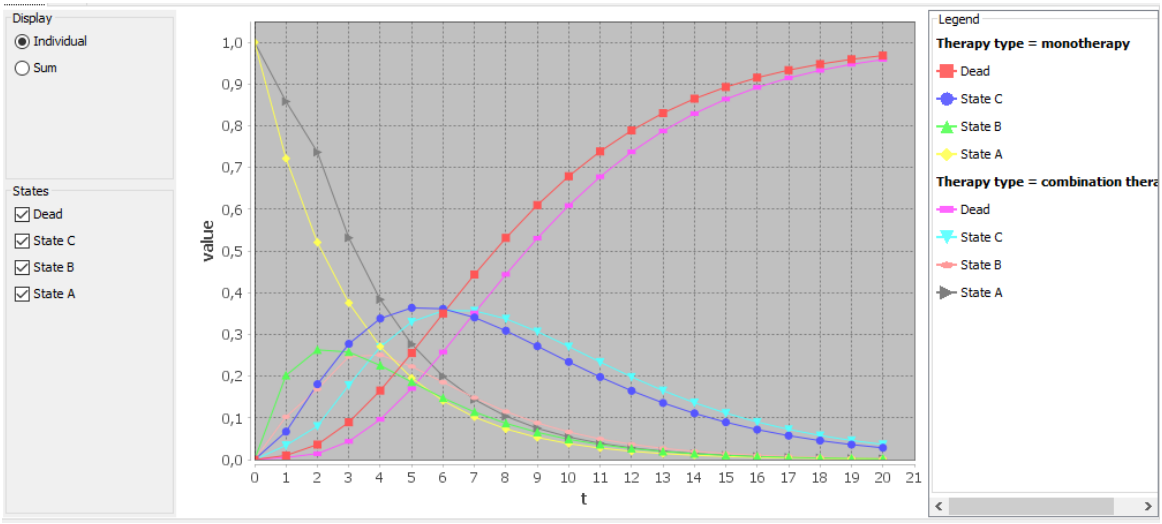


Figure 5.6: Temporal evolution of the variable *State [i]*.

slice.

Algorithm 5.1 can be used for three different purposes:

1. Performing CEA efficiently, i.e. without doing inference with all the potentials at the same time it consumes less memory than other algorithms.
2. Applying the numerical integration techniques explained in Chapter. 4 using the values of cost and effectiveness of each cycle
3. Displaying the temporal evolution of a variable of interest, i.e. showing either the value of a numeric variable or the probability of each state of a finite-states variable. In Figure 5.6 we can see the temporal evolution of the variable *State [i]* of the MID for the HIV model of Chancellor et al. (1997) (see Fig. 2.6).

5.3 Cost-effectiveness analysis with decision analysis networks

As we said in Section 2.3.4, DANs are a new type of PGM specially designed for asymmetric problems.

To solve CEA problems with DANs, we combined the method for unicriterion DANs and the CEA algorithms for DTs with embedded decision nodes (Arias and Díez, 2011) and for IDs (Arias and Díez, 2015). The basic idea is straightforward, but

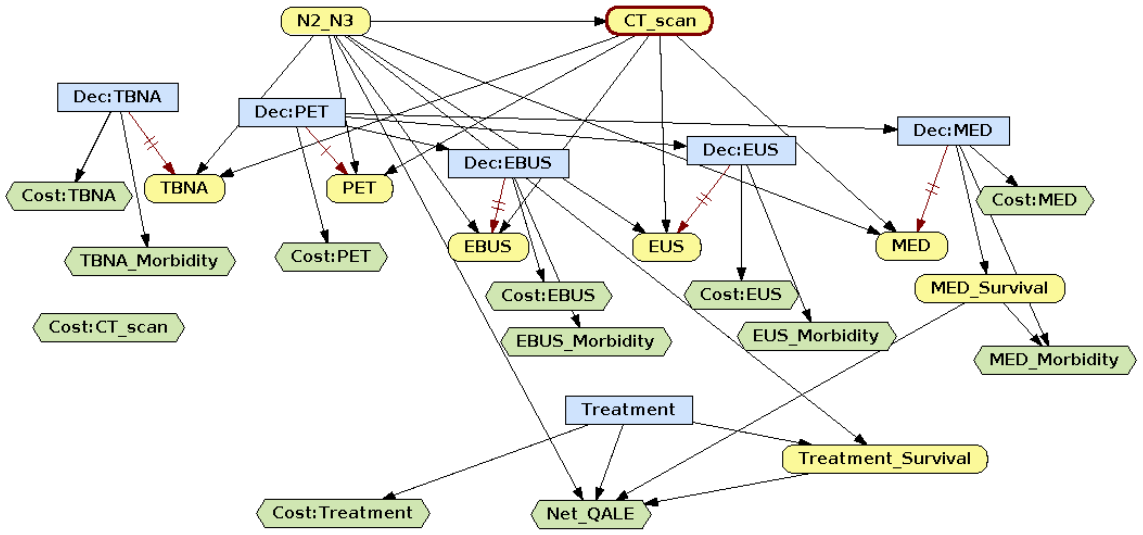


Figure 5.7: Mediastinet, a DAN for the mediastinal staging of non-small cell lung cancer.

the integration of those algorithms is not trivial.

We illustrate the method with some examples. The first one is a fictitious problem involving two mutually-exclusive therapies and n tests. The goal is to find, for each value of λ, the optimal intervention, i.e., to determine which test must be done first, if any, and depending on its result, whether to do a second test, etc., and finally to decide what therapy to apply, if any. We solve in detail a numerical example, with two tests ($n = 2$) and examine empirically the computational time required to solve this problem for different values of n. We then show how to solve a real-world problem: finding the optimal sequence of six tests available for the mediastinal staging of non-small cell lung cancer (see Fig. 5.7). We solved this problem with an ID (Luque et al., 2016), which we have now converted into a DAN.

We introduce here the n-test problem as an example involving unordered decisions. In Section 5.3.1 we discuss why, in spite of its apparent simplicity, it is difficult to determine the optimal intervention—as a function of λ—even for only two tests ($n = 2$). Then we will show how to solve the problem using DANs.

Example 5.1. (The n-test problem). For a disease X There are two mutually-exclusive therapies and n tests, each having two possible outcomes, positive and negative. Every test can be performed once at most.

In order to solve the problem numerically, we will assume that are only two

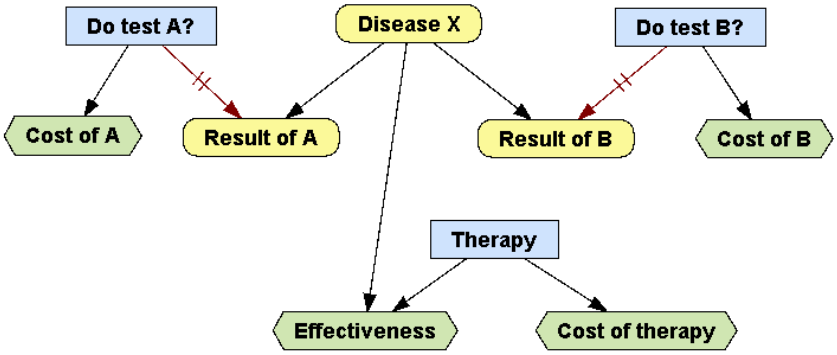


Figure 5.8: A DAN for the 2-test problem.

Therapy	Cost	Effectiveness	
		disease	**no disease**
no therapy	€0	1.2 QALY	10.0 QALY
therapy 1	€20,000	4.0 QALY	9.9 QALY
therapy 2	€70,000	6.5 QALY	9.3 QALY

Table 5.1: Cost and effectiveness of each intervention for the 2-test example.

tests, A and B, such that the latter is more accurate, but more expensive, as shown in Table 5.2. The DAN for this problem is shown in Figure 5.8.

We will suppose that the prevalence of X is 0.14. The effectiveness of the therapies depends on whether the disease is present or not, as shown in Table 5.1. The best situation for the patient occurs when the disease is absent and no therapy is applied (effectiveness = 10.0 QALY). The worst situation occurs when the disease is present and no therapy is applied (effectiveness = 1.2 QALY). For sick patients, the second therapy yields 6.5 QALY while therapy 1 only yields 4.0, but when a therapy is applied to healthy patients (by mistake) it reduces the effectiveness due to side effects, especially therapy 2.

Test	Cost	Sensitivity	Specificity
A	€18	78%	91%
B	€150	90%	93%

Table 5.2: Cost, sensitivity and specificity of each intervention for the 2-test example.

5.3.1 Complexity of modeling the n-test problem with decision trees

We have introduced in Example 5.1 the n-test problem. We analyze here the complexity of modeling it with DTs, first for the unicriterion case and then for CEA.

5.3.1.1 Decision trees for unicriterion analysis

If cost and effectiveness were combined into a single criterion, the NMB (cf. Eq. 3.2), the optimal intervention could be found by building the DT. When there is no test ($n = 0$), the tree contains the decision node "Therapy" at the root, with a "Disease X" node in each branch. If we denote by $l(n)$ the number of leaves in the tree for n tests, we have $l(0) = 3 \times 2 = 6$. When there is one test, the decision node at the root has two branches: "do the test" and "do no test". The first one has two branches, "positive" and "negative". Each of these three branches has a "Therapy" node, as in the previous case, so the tree has $l(1) = 3 \times 3 \times 2 = 18$ leaves. When there are n tests, the root has one branch, "do no test", and n branches "do test A", "do test B", etc. The "no test" branch has 6 leaves; each of the other branches has $l(n-1)$ leaves for a positive result of the test and $l(n-1)$ leaves for a negative result. Therefore, $l(n) = 2n \cdot l(n-1) + 6$, and we have $l(2) = 78$, $l(3) = 474$, $l(4) = 3,798$, etc. Consequently the problem is hard for two tests and virtually impossible to solve for more than two tests. If there were more than two therapies and each test had more than two outcomes, the problem would be much harder even for only two tests.

5.3.1.2 Decision trees for cost-effectiveness analysis

The standard roll-back algorithm for CEA evaluates DTs by assigning a cost-effectiveness pair to each node, starting from the leaves, and then performing a CEA at the root, which returns a set of ICERs. This algorithm cannot be applied to the trees described in the previous section (except for $n = 0$) because the evaluation of an embedded decision node does not return a single cost-effectiveness pair.[1] The

[1] Old versions of TreeAge evaluated embedded decision nodes by asking the user for a value of λ in order to select for each node the branch that maximizes the NMB. The question might disconcert the user, because it is not necessary: the tree must be evaluated for all the values of λ,not for a particular one. Recent versions use a default value, which can be set by the user. Using a single value of λ makes it is possible to assign a cost-effectiveness pair to each node, but the evaluation

solution proposed by Kuntz and Weinstein (2001) consists in building a DT containing just one decision node, at the root, with an outgoing branch for each intervention. When no test is available ($n = 0$), there are only three possible interventions: "no therapy", "therapy 1", and "therapy 2". Each branch has a chance node, X, with two outgoing branches, "present" and "absent". Therefore, in this case the tree has 6 leaves, $l(0) = 6$.

When $n = 1$, there are three no-test interventions, as in the previous case, plus each with a plus 3×3 do-test interventions: "if test is negative, then no therapy; if positive, no therapy", "if test is negative, no therapy; if positive, therapy 1", "if test is negative, no therapy; if positive, therapy 2", "if test is negative, then therapy 1; if positive, therapy 2", ... , "if positive, therapy 2; if negative, therapy 2". The 3 interventions in which the therapy applied is the same regardless of the result of the test can be ruled out because it is not worth doing a test that will not guide the decision about the therapy. Therefore, there are 3×2 do-test interventions. Each one has 2×2 leaves, because the chance node that represents the result of the test has two outgoing branches, "positive" and "negative", and node for X has two outgoing branches, "present and absent". Therefore, the tree has $6 + 6 \times 4$ leaves, i.e., $l(1) = 30$.

When $n = 2$, there are 3 no-test interventions; each one corresponds to a branch outgoing from the root of the tree, and has two leaves. There are also some interventions that begin doing test A. When it is positive, there are 3 subsequent interventions that do not perform test B (as many as when there was not test available) and 6 interventions that perform it (as many as if there were only one test available), i.e., 9 sub-interventions. There are also 9 sub-interventions for a negative result of test A. This makes a total of 9×9 interventions that begin doing test A. However, it is not worth doing the test when it does entail a difference in the sub-intervention applied after it, so we can limit our analysis to $9 \times 8 = 72$ interventions. Each branch for an intervention in which both tests are done contain $2 \times 2 \times 2 = 8$ leaves. An intervention that performs only test A leads to $2 \times 2 = 4$ leaves. When test B is done only for one of the outcomes of test A, the branch contains 6 leaves. This makes a total of 480 leaves for the interventions that begin doing test A. There are also 72 interventions that begin doing test B. This makes a total of $3 + 72 + 72 = 147$ interventions and

depends on the value arbitrarily chosen and the result is often wrong (Arias and Díez, 2014).

$6 + 480 + 480 = 966$ leaves.

For $n \geq 3$ the possible interventions are so complex that it is very difficult even to estimate their number, but the increase from $l(1) = 30$ to $l(2) = 966$ leaves shows the impressive increase of the size of the tree for every test we add.

We have excluded from these counts the interventions in which the result of a test does not guide the next decisions—for example, the interventions that apply a therapy regardless of the result of the test. Depending on the numerical parameters of the model we might discard more interventions. For example, in the one-test problem we can see that the incremental effectiveness of "therapy 1" with respect to "no therapy" grows with the probability of disease X. Therefore, the intervention "when test A is negative apply therapy 1; when positive, apply no therapy" is clearly suboptimal. However, when the number of decisions increases, the proportion of therapies that can be discarded is smaller and smaller.

In summary, building manually a DT with only one decision node for CEA, as proposed by (Kuntz and Weinstein, 2001), is unfeasible for the 2-test problem and absolutely impossible for $n > 2$.

5.3.2 Evaluation of the DAN

Cost-effectiveness DANs can be evaluated with Algorithm 5.2, which recursively decomposes the original network into DANs without decision nodes. We illustrate its performance for 2-test DAN (cf. Fig. 5.8). We denote the variables by capital letters and their values (states) by the corresponding lowercase letters. Thus, X represents the disease and $+x$ and $\neg x$ mean that it is present or absent respectively, R_A represents the outcome of test A and $+r_A/\neg r_A$ a positive/negative result, etc.

The algorithm is first invoked with the original DAN and no evidence, because no variable has yet been observed. The first decomposition generates two new DANs, as shown in Figure 5.9; in one of them the first decision is whether to do test A or not, and in the other the first decision is whether to do test B. The decomposition of the former generates two new DANs, in which D_A disappears because every decision node is deleted when making the decision; in one of these DANs the option chosen

is to do the test, $+d_A$, which reveals the value of R_A; in the other the option chosen is not to do the test, $\neg d_A$, which is incompatible with the two values of R_A (because in this case the test is neither positive nor negative), so node R_A does not appear in this network. The DAN for $+d_A$ is then decomposed into two new DANs: one for the positive result of the test ($+r_A$) and another one for a negative result ($\neg r_A$). The process continues until the network contains no decisions and every observable variable has been assigned a value. In this case the algorithm returns a probability (a single number) and a CEP, and the recursion continues backwards. At the end of the process the probability returned is 1 and the CEP contains the optimal intervention for each value of λ. The process is explained in more detail in Appendix B.

Every node in the decomposition tree (Fig. 5.9) corresponds to a node in a DT (Fig. 5.10), which implies that Algorithm 5.2 can be adapted to generate a DT if we wish, provided that our computer has enough memory to store it.

5.3.3 Results

Figure 5.11 shows the conclusion of the CEA for the 2-test DAN, which solves the numeric problem stated in Example 5.1. The CEP consists of 6 intervals, each having an optimal interventions. When the WTP is very low, i.e., $\lambda <$ €7.718.95/QALY, it is not worth doing either test because, regardless of its result, no therapy will be applied. For all the other intervals, a therapy is applied only when the last test performed is positive. In the second interval, i.e., €7,718.95/QALY $< \lambda <$ €21,385.50/QALY, test A should be done; if it is positive, the first therapy should be applied. The optimal intervention for the third interval is similar, with therapy 2 instead of therapy 1. In the fourth interval test B should be done first; if it is positive, test A must be done to determine the optimal treatment: therapy 1 when A is negative and therapy 2 when it is positive. In the fifth interval the optimal intervention is to do both tests, in any order, even if the first one is negative; if one of them is positive, therapy 1 must be applied; if both are positive, the most beneficial treatment is therapy 2. In the sixth interval, the WTP exceeds €113.139.00/QALY; a positive result in test B leads to the direct application of therapy 2, while a negative result requires doing test A and applying therapy 1 when this is positive.

Algorithm 5.2: Cost-effectiveness analysis for a DAN

function : `evaluateDAN`
input : DAN – a decision analysis network,
$\mathbf{e}$ – evidence (a configuration of a set of variables $\mathbf{E}$)
output : $P(\mathbf{e})$ – the probability of the evidence,
$CEP(\mathbf{e})$ – a cost-effectiveness partition

if *DAN has decision nodes* **then**
  $\mathbf{O} \leftarrow$ always-observed variables in DAN, except those in $\mathbf{E}$
else
  $\mathbf{O} \leftarrow$ all the chance variables in DAN, except those in $\mathbf{E}$
if $\mathbf{O} = \varnothing$ *and DAN has no decisions* **then**
  $P(\mathbf{e}) \leftarrow$ product of the projected probability potentials
  $CEP(\mathbf{e}) \leftarrow$ sum of the projected utility potentials
else if $\mathbf{O} \neq \varnothing$ **then**
  select $X \in \mathbf{O}$ such that X has no ancestor in $\mathbf{O}$
  foreach x *of* X **do**
    $DAN_x \leftarrow$ `instantiate` (DAN, X, x)
    $\{P(\mathbf{e} \circ x), CEP(\mathbf{e} \circ x)\} \leftarrow$ `evaluateDAN(`DAN_x, $\mathbf{e} \circ x$`)`
  $P(\mathbf{e}) \leftarrow \sum_x P(\mathbf{e} \circ x)$
  $P(x \mid \mathbf{e}) \leftarrow P(\mathbf{e} \circ x)/P(\mathbf{e})$
  $CEP(\mathbf{e}) \leftarrow$ `averageCEP`$(X, P(x \mid \mathbf{e}), \{CEP(\mathbf{e} \circ x_1), \ldots, CEP(\mathbf{e} \circ x_m)\})$
else if *exactly one decision* D *can be made first* **then**
  foreach d *of* D **do**
    $DAN_d \leftarrow$ `instantiate` (DAN, D, d)
    $\{P_d(\mathbf{e}), CEP_d(\mathbf{e})\} \leftarrow$ `evaluateDAN(`DAN_d, $\mathbf{e}$`)`
  $P(\mathbf{e}) \leftarrow P_d(\mathbf{e})$ for an arbitrary value d
  $CEP(\mathbf{e}) \leftarrow$ `optimalCEP`$(D, \{CEP_{d_1}(\mathbf{e}), \ldots, CEP_{d_m}(\mathbf{e})\})$
else
  $\mathbf{D}_I \leftarrow$ *decisions that can be made first*
  foreach D *in* $\mathbf{D}_I$ **do**
    $DAN_D \leftarrow$ `prioritize` (DAN, D)
    $\{P_D(\mathbf{e}), CEP_D(\mathbf{e})\} \leftarrow$ `evaluateDAN(`DAN_D, $\mathbf{e}$`)`
  $P(\mathbf{e}) \leftarrow P_D(\mathbf{e})$ for an arbitrary decision D
  $OD \leftarrow$ new decision variable (meta-decision), with one value for each decision D in $\mathbf{D}_I$
  $CEP(\mathbf{e}) \leftarrow$ `optimalCEP`$(OD, \{CEP_{D_1}(\mathbf{e}), \ldots, CEP_{D_m}(\mathbf{e})\})$
return $\{P(\mathbf{e}), CEP(\mathbf{e})\}$

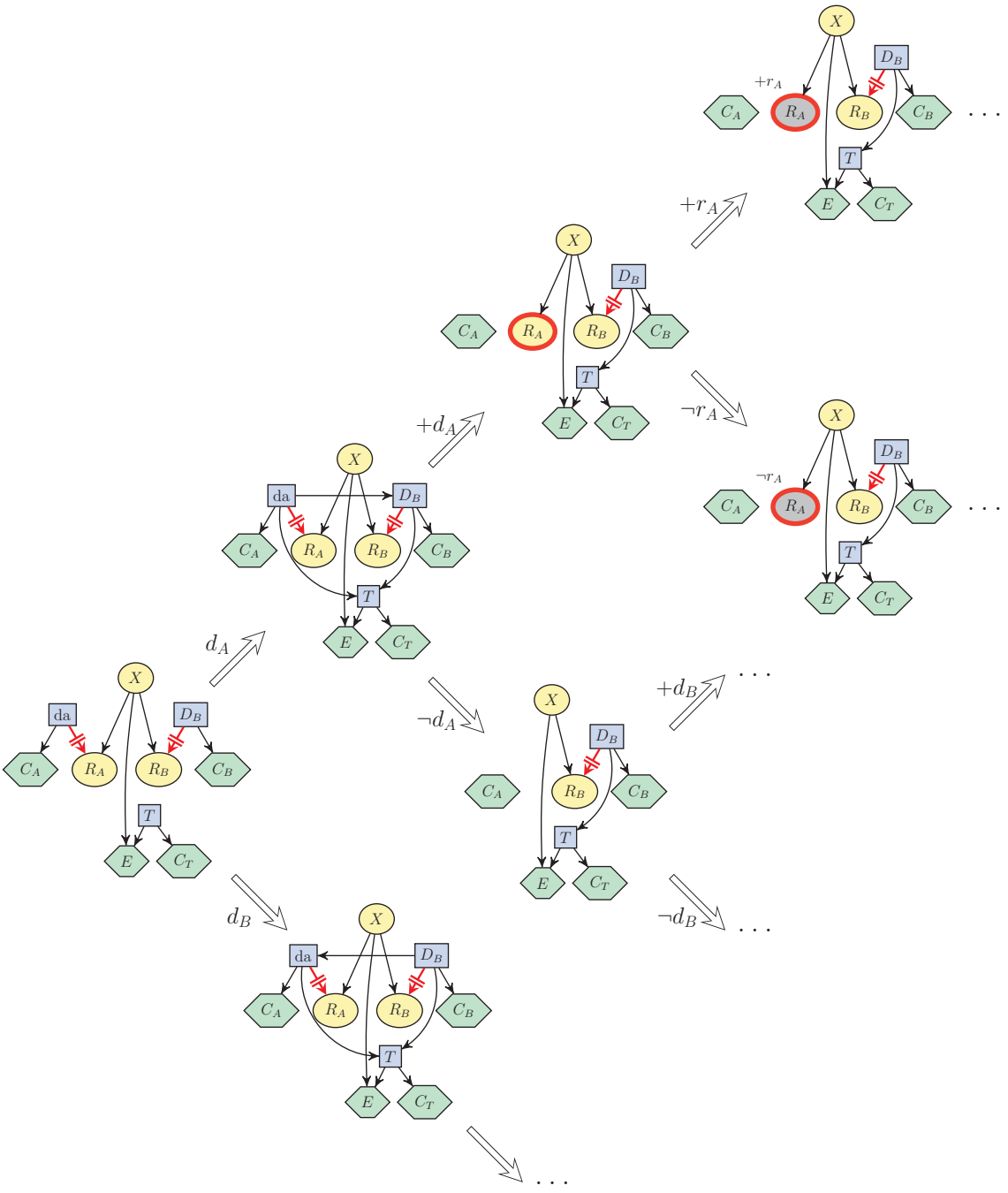


Figure 5.9: Recursive decomposition of the DAN for the 2-test problem. Every node in this tree corresponds to an invocation of evaluateDAN (Algorithm 5.2). In the DANs, chance nodes having associated evidence are colored in gray.

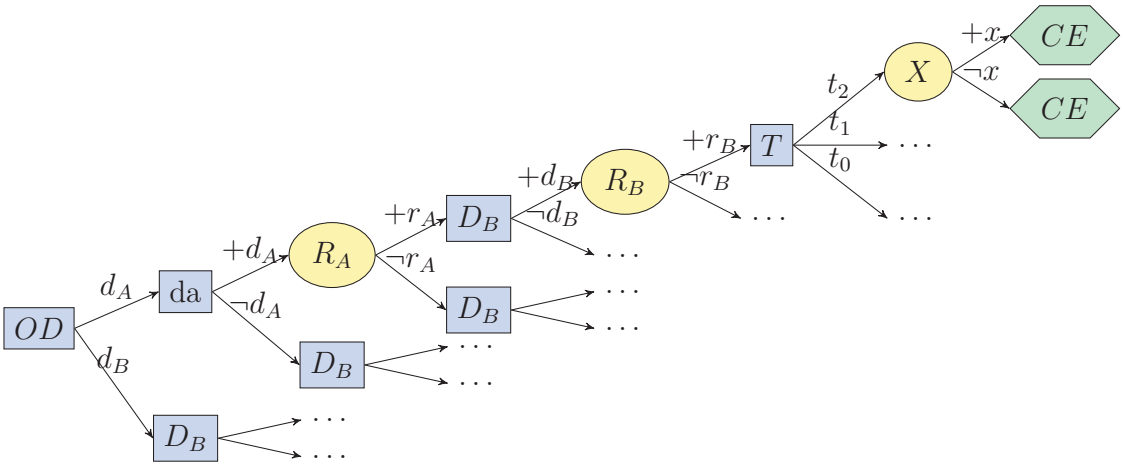


Figure 5.10: DT generated from the DAN for the 2-test problem. Every node in this tree corresponds to a node in the decomposition tree (Fig. 5.9), i.e., to an invocation of evaluateDAN. The "meta-decision" node *OD* (order of the decisions) determines which decision to make first.

λ inf.	λ sup.	Cost	Effectiveness	Intervention
0.0	7718.95	0.0	8.768	OD = Do test A? -> Do test A? = no -> Do test B? = no -> Therapy = no therapy
7718.95	21385.5	2119.95	9.04264	OD = Do test A? -> Do test A? = yes -> IF Result of A = negative -> Do test B? = no -> ...
21385.5	24361.7	7304.85	9.28509	OD = Do test A? -> Do test A? = yes -> IF Result of A = negative -> Do test B? = no -> ...
24361.7	71550.3	9062.25	9.35723	OD = Do test B? -> Do test B? = yes -> IF Result of B = negative -> Do test A? = no -> ...
71550.3	113139.0	10734.9	9.38061	OD = Do test A? -> Do test A? = yes -> IF Result of A = negative -> Do test B? = yes -> ...
113139.0	+∞	14856.7	9.41704	OD = Do test B? -> Do test B? = yes -> IF Result of B = negative -> Do test A? = yes -> ...

Figure 5.11: CEP for the 2-test problem. It is the final result of evaluating the DAN in Figure 5.8 with Algorithm 5.2.

It is possible to solve the n-test problem for a higher number of tests. Adding a new test to the DAN takes less than 2 minutes in OpenMarkov's graphical user interface, which is negligible compared to the time required to add a new test in a DT. On a desktop computer, Algorithm 5.2 can evaluate the DAN for $n \leq 3$ in less than a second, for $n = 4$ in 7 seconds, for $n = 5$ in 2 minutes, for $n = 6$ in 40 minutes, and for $n = 7$ in 17.5 hours.

Algorithm 5.2 is an adaptation for CEA of Algorithm 1 in (Díez et al., 2018a). The CEA version of Algorithm 2 in that paper, which is much more efficient, can solve the case $n = 7$ in 9 minutes and $n = 8$ in 3 hours. With parallel computation the evaluation would be much faster. These are remarkable results because the DTs (with embedded decision nodes) have more than 6 million leaves for $n = 7$ and more than 100 million for $n = 8$—see Section 5.3.1.

The DAN Mediastinet, which contains 5 partially-ordered tests, can be evaluated

in less than 3 seconds. The resulting CEP contains 18 intervals. Figure 5.12 shows the optimal intervention for $\lambda = €30{,}000$/QALY, which is the shadow cost-effectiveness threshold estimated for the Spanish public health system (Sacristán et al., 2002; de Cock et al., 2007). This intervention differs from the one we obtained from the ID that used the same parameters, which applies chemotherapy when the endobronchial ultrasound test (EBUS) is positive—see Figure 5 in (Luque et al., 2016)—while the DAN recommends doing transesophageal ultrasound-guided fine needle aspiration (EUS) and mediastinoscopy (MED) in this case.

5.3.4 Discussion

Using DANs it is possible to solve many decision problems that cannot be modeled with traditional techniques. One reason is that the standard roll-back algorithm for CEA cannot evaluate DTs containing embedded decision nodes (Arias and Díez, 2014). It would be necessary to build a tree containing one decision node, at the root, with an outgoing branch for each intervention, as proposed by Kuntz and Weinstein (2001). In Subsection 5.3.1 we show that when no test is available ($n = 0$), there are 3 intervention, one for each therapy, so the tree has only 6 leaves. When $n = 1$ there are 9 interventions—after excluding some that are clearly suboptimal—and the tree has 30 leaves. When $n = 2$ there are 147 interventions and the tree grows up to 966 leaves. For $n \geq 3$ the interventions are so complex that it is very difficult to estimate their number, but the increase from 30 to 966 leaves shows us the impressive growth of the tree for every test we add.

Due to the complexity of the problem, economic evaluations involving unordered decisions use expert knowledge either to impose a total ordering among them or to select a very small number of interventions. For example, a CEA of four tests for the diagnosis of coronary heart disease only examined 8 out of the thousands of possible interventions (Walker et al., 2013). It is likely that the shortlist contained the optimal intervention, but it is not necessarily so.

In fact, if we reexamine the six interventions resulting from the evaluation of the 2-test problem, which we presented in Section 5.3.3, we can see that the first three ones are straightforward, but those for the fourth and the last intervals are not so obvious.

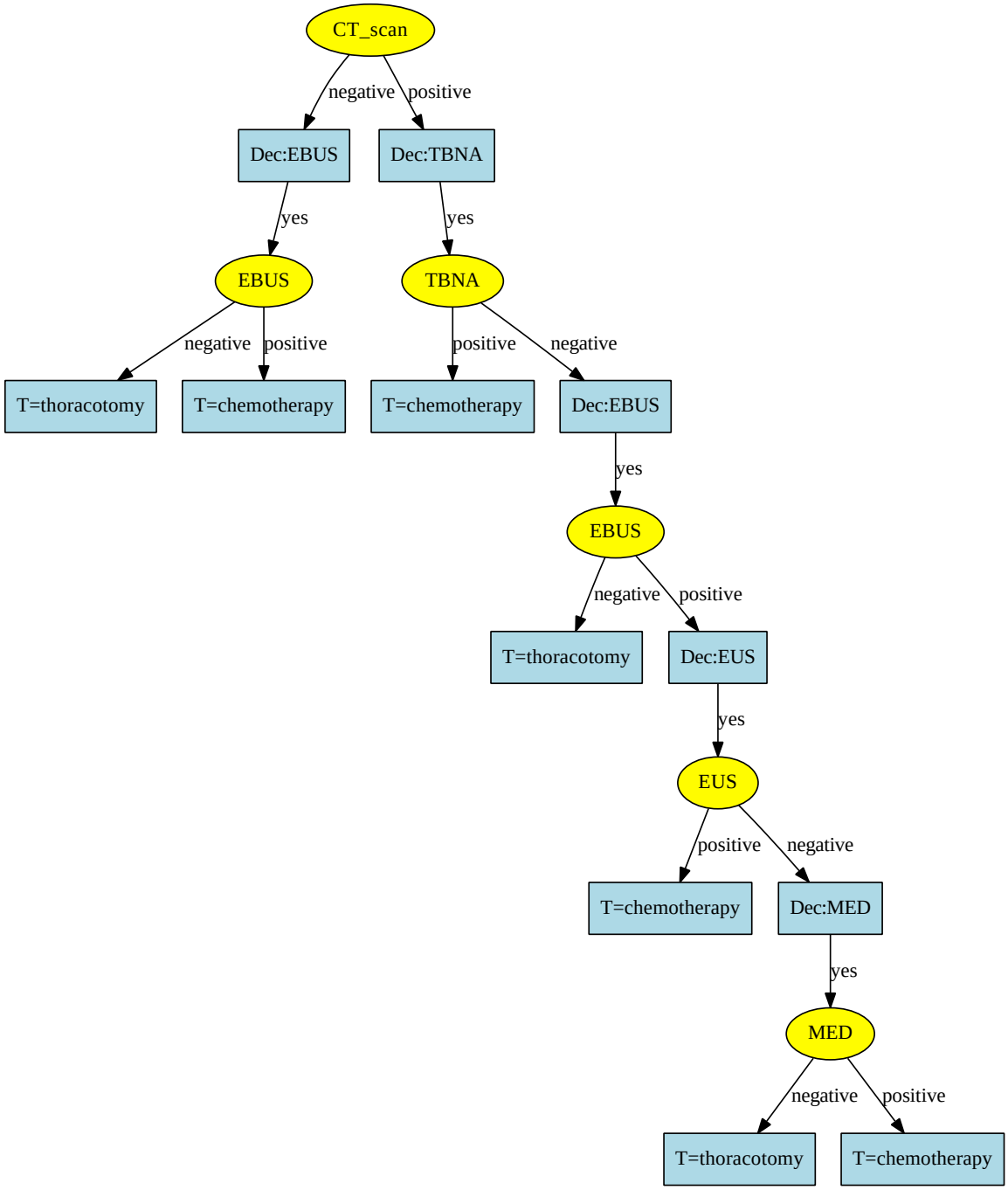


Figure 5.12: Optimal intervention for $\lambda = €30{,}000$/QALY obtained from the DAN Mediastinet (cf. Fig. 5.7).

Some decision analysts, including ourselves, might have overlooked the intervention that turned out to be optimal for $\lambda >$ €113.139.00/QALY: "do test B first; if positive, apply therapy 2; if negative, do test A and if this is positive apply therapy 1". That strategy is counterintuitive, because apparently no therapy should be applied when test B, which is more sensitive and specific than A, has ruled out the presence of the disease. However, a numerical analysis shows that $P(+x| + r_A, \neg r_B) = 0.1317$, and for this posterior probability therapy 1 is more effective than both no therapy and therapy 2. This situation reminds us of the application of AI to games, such as go and chess: even the best human players examine only a very small amount of combinations, while computers, which evaluate many more possibilities, sometimes find excellent moves that no human would have figured out. Similarly, when we built the ID Mediastinet, the pneumologist who helped us did not know whether EBUS should be done before EUS or vice versa. This questions has been solved with the DAN version of the same model (cf. Fig. 5.12).

DANs are not only superior to the traditional way of doing CEA with DTs. They also outperform the methods we have proposed previously. One of these is the algorithm for evaluating cost-effectiveness DTs with embedded decision nodes (Arias and Díez, 2011). These trees do not need one branch for each intervention, because their evaluation implicitly compares all the possible interventions. So in the n-test problem they only need 18 leaves (instead of 30) for $n = 1$ and 78 leaves (instead of 966) for $n = 2$. However, it is still impossible to manually build the trees for a higher number of tests, which require 474 leaves for $n = 3$, 2,845 leaves for $n = 4$, etc. By contrast, DANs can evaluate up to the case $n = 8$, for which the DT would contain more than 100 million branches, as mentioned above.

DANs are also superior to IDs. These can perform CEA for large problems (Arias and Díez, 2015), but only if the decisions are totally ordered. So when building an ID for the 2-test problem we must decide which test to do first, either A or B, and it would therefore be impossible to obtain the CEP shown in Figure 5.11.

Additionally, the use of restrictions allows DANs to usually avoid dummy states, such as "test not done", which in an ID must be added to the outcomes of a test to indicate that when we decide not to do the test the result is neither "positive" nor "negative". Dummy states unnecessarily increase the size of the probability tables and

the number of parameters, some of which are 0 or 1 (Bielza et al., 2010; Díez et al., 2018a). For this reason even when a problem can be modeled with an ID, the DAN representation is cleaner and more elegant.

The main drawback of DANs is that health economists, who are all familiar with DTs, would need to learn a new formalism. However, the effort is small for someone acquainted with the fundamentals of decision analysis, and the possibility of easily solving much larger problems much more easily will soon compensate for the time invested. Additionally, the availability of an open-source tool for DANs will avoid the need to acquire software licenses for building and evaluating DTs. In fact, OpenMarkov can automatically convert a DAN into a DT; if the tree is too big, it can be expanded only to the desired depth. This implies that the decision modeler who builds a DT only has a DT, while the one who builds a DAN has both the DAN and the tree, with much less effort.

We conclude this section with a comment about efficiency. Using DANs it is possible to model problems involving many decisions. Adding a new test to a decision problem takes less than two minutes in OpenMarkov, so we can model the n-test problem for arbitrary large values of n. The problem lies in the time complexity, which in general increases much faster than exponentially with the number of nodes in the DAN. Fortunately, Algorithm 5.2, which is equivalent to the evaluation of a DT, can easily be implemented in parallel, using a negligible amount of memory for each branch. Alternatively, it is possible to use a more efficient decomposition of the DAN (cf. Algorithm 2 in (Díez et al., 2018a)), which can take profit of tensor-processing libraries that run on GPUs and TPUs.

Part IV

Medical applications

6

Cost-effectiveness of bilateral cochlear implantation

"For people without disability, technology makes things easier. For people with disabilities, technology makes things possible."

Mary Pat Radabaugh

Results of this work were presented in Pérez-Martín and Díez (2015) and Pérez-Martín et al. (2016) and later published in the journal *The Laryngoscope* (Pérez-Martín et al., 2017a)..

6.1 Introduction

6.1.1 Audition and hearing loss

According to the World Health Organization (WHO) around 466 million people have disabling hearing loss, and 34 million of these are children[1]. They have estimated that untreated hearing loss involves an annual global cost of US$ 750 billion, which makes hearing loss one of the most common chronic diseases around the world. This decrease in hearing ability is usually linked to age (one third of people over 65 years of age are affected by disabling hearing loss), but also affects younger sectors of the population.

Deafness significantly influences an individual's social and learning development, especially if it occurs prior to the development of lingual capacity (pre-speech deafness). An individual with deafness will not only require a greater effort of adaptation in their daily life, but will also need additional reinforcement elements, such as the learning of sign language or warning systems to identify some sounds.

There are several causes of hearing loss: infections, traumatisms, hereditary hearing disorders, autoimmune diseases, adverse drug reactions, circulatory problems, neurological diseases (such as multiple sclerosis), etc. (Sataloff and Sataloff, 2005). Sensorineural deafness is the most common cause of permanent hearing loss and it was usually caused by a damage in the inner structure of the cochlea.

The cochlea is the part of the inner ear involved in the hearing process. It is a spiral-shaped cavity in the bony labyrinth, which is formed by the cochlea, the vestibule, and three semicircular canals. The cochlea is divided in three tubes: the scala vestibuli, the scala tympani and the cochlear duct. The scala tympani and the cochlear duct are separated by the basilar membrane, which it lodges the organ of Corti, the main responsible of the perception of sound. It is formed by 3 ducts that

[1] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

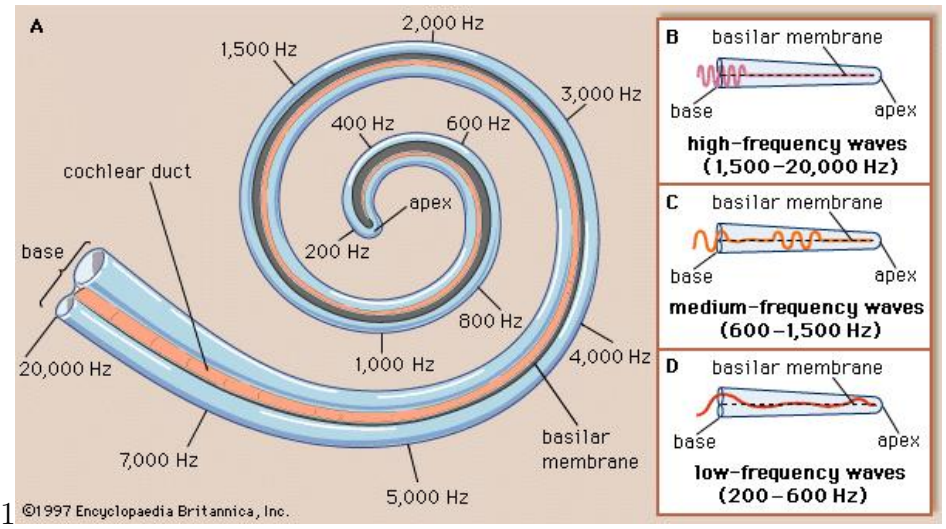
6.1 

Figure 6.1: The analysis of sound frequencies at the basilar membrane. (A) The fibers of the basilar membrane become progressively wider and more flexible from the base of the cochlea to the apex. As a result, each area of the basilar membrane vibrates preferentially to a particular sound frequency. (B) High-frequency sound waves cause maximum vibration of the area of the basilar membrane nearest to the base of the cochlea; (C) medium-frequency waves affect the centre of the membrane; (D) and low-frequency waves preferentially stimulate the apex of the basilar membrane.Encyclopaedia Britannica ,Inc.

contain the hair cells, which transform the sound vibration (a mechanical wave) into a neural signal (an electrical signal) that can be processed by the brain. When the pressure of the sound moves the fluids of the inner ear, the basilar membrane deforms in a specific area, depending on the frequency of the sound: low frequencies stimulate the hair cells near the apex, while higher frequencies stimulate those close to the base, as shown in Figure 6.1.

There are several techniques and technologies for the partial recovery of hearing. However, their effectiveness varies depending on the etiology of deafness.

Hearing aids are external devices that amplify the sound. They are relatively cheap and very effective for low to mild deafness.

However, for severe to profound sensorineural hearing loss, Cochlear Implants (CIs) are much more effective (Clark, 2003; Papsin and Gordon, 2008). The loss of hearing ability due to damage to the cochlea also makes it difficult to understand the sounds, and can be perceived in a distorted or muffled way. It can also cause different effects depending on the intensity or frequency of the sound, being able to clearly hear

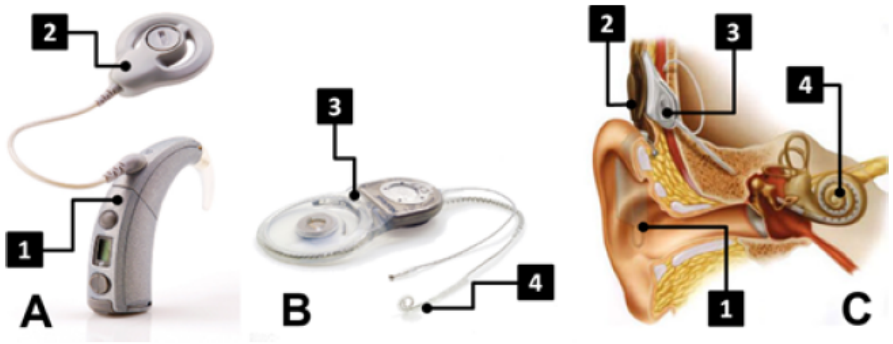


Figure 6.2: Cochlear implant components. The CI (C) has two main components, one external (A) and the other internal (B). External component is composed by the microphone and the speech processor (1) and the coil (2). Internal component is composed by the receiver-decoder-stimulator (3) and the electrode array (4).

some sounds and not others (of greater or lesser frequency or intensity).

6.1.2 Cochlear implants

A CI is an electronic medical device that allows the recovery of audition in cases of severe to profound hearing loss. Its implantation requires a surgical operation in which the electrode array is placed along the spiral of the cochlea. Each electrode in the array is responsible to stimulate a set of cells of the auditory nerve and, therefore, will be responsible for a specific frequency band (see Fig. 6.1).

As shown in Figure 6.2, it has two main components, one external and one internal. The external component captures the sound of the environment with a microphone and converts it into a digital signal, that is transmitted from the external coil to the internal coil, placed under the skin. The internal decoder transforms the signal into electrical impulses, which are sent to the electrode array placed inside the cochlea to stimulate the hair cells, the hearing nerves and finally the neurons in the brain.

The first multichannel CI—with more than one electrode in the array—was implanted in 1976 at the Saint-Antoine hospital in Paris (Chouard et al., 1977). In 1984 the Food and Drug Administration (FDA) of the United States authorized the CI in adults and in 1990 in children. The first BCI (that is, a device implanted in

each ear) was made in 1996[2]. Since then BCI has become more and more frequent, although with a very unequal use in each country.

Children who have only one CI have greater difficulty understanding language (even with mild noise), and the effort of concentration necessary makes them feel tired at the end of the day (Summerfield et al., 2003). In contrast, BCI significantly increases the understanding of oral language in children and adults, especially in group conversations and in noisy environments (Peters, 2007). In addition, the BCI allows users to identify the source of the sound, improving understanding of spoken language (especially in noisy environments), allowing continue hearing when the first implant fails, increasing traffic safety (when crossing a road, riding a bicycle, driving a car, etc.) (Johnston et al., 2009; Sparreboom et al., 2010; Crathorne et al., 2012; Schoonhoven et al., 2013). It also improves the integration of children in the classroom and the school performance: In 1986 it was shown that children with unilateral hearing loss were 10 times more likely to discontinue a course (Bess, 1986); in Belgium, it was found in 2007 that all the children with BCI studied in ordinary schools, while 46% of those who carried a unilateral implant attended specific schools for deaf people(Raeve, 2007). On the other hand, the BCI allows to continue listening when one of the devices run out of battery, when one of the two processors is damaged or when the internal part of the implant fails. In some cases it has been necessary to remove the internal part of the implant and it has not been possible to reimplant it in the same ear; if the other ear has not been developed because it has never received acoustic stimulation, the second implant would then have a very low efficiency. For all these reasons, the experts, mainly in the United States and the United Kingdom, recommend without a doubt the pediatric BCI (Offeciers et al., 2005; Peters, 2007; Balkany et al., 2008; Papsin and Gordon, 2008) and some of them even consider that it could be cost-effective for certain groups of adults.

Although BCI has several advantages, cost-effectiveness studies have not proved unequivocally the efficiency of the BCI.

[2] http://www.medel.com/about-Med-el/

6.1.3 Previous cost-effectiveness studies

The first study that measured the increment in quality of life from unilateral to bilateral CI was conducted by Summerfield et al. (2003). The value obtained, 0.3 was used in almost all the CEAs done in the next years. Using the quality of life estimated at a previous study (Summerfield et al., 2002) they determined that the ICER of BCI with respect to UCI for adult people was of £100,000/QALY, much higher than the WTP established by the NICE, which at this time was of £50,000/QALY.

In 2006 the NICE commissioned the elaboration of a clinical guideline for cochlear implantation. The manufacturers of CIs were invited to submit their CEAs of the BCI. Cochlear Europe Ltd. presented one, cited in (Bond et al., 2007), which estimated an ICER of £32,909/QALY for adults and £39,049/QALY for children. This is counterintuitive because the benefits of BCI are higher for children than for adults, and hence the ICER should be lower.

At the end of that year Barton et al. (2006) showed that simultaneous BCI (implanting both devices in the same surgical operation) is cost-effective for certain groups with severe to profound hearing loss that do not benefit from hearing aids: children with prelocutive deafness; adults and children that are also blind; adults and children with risk of ossification of the cochlea (which usually occurs shortly after meningitis). Due the weakness of the arguments in that study, NICE published in March 2008 a second version of the guideline, which recommended BCI only for children and adults with blindness or at risk of ossification of the cochlea—see (Raine et al., 2010) for more information about this publication and the subsequent controversy.

In 2007, the Agencia para la Formación, Investigación y Estudios Sanitarios de la Comunidad de Madrid Pedro Laín Entralgo, in Spain, made a brief CEA of pediatric BCI (L-Pedraza Gómez et al., 2007), also based on the gain in quality of life obtained by Summerfield et al. (2002). Costs were obtained from Spanish institutions. The study concluded that the ICER was €53,018/QALY for adults in case of simultaneous BCI and €63,487/QALY in case of sequential implantation (implanting both devices in two different surgical operation). For children, the ICERs were €44,199/QALY and €56,640/QALY, respectively. As mentioned in Section 3.1.2, the WTP for the Spanish public health system, accepted as a consensus among experts, is around

€30,000/QALY, implied that BCI was not cost-effective in our country.

Shortly afterwards, Bichey and Miyamoto (2008) published a CEA of the BCI based on their own measurements of the increase in quality of life, obtaining much lower ICERs than in other studies, as shown in Table 6.1. They used the HUI instead of the quality of life estimate of (Summerfield et al., 2002). Surprisingly, in their study, the increment in quality of life from UCI to BCI is higher than from profound deafness to UCI.

Two years later Summerfield et al. (2010) published a CEA of the BCI based on a Markov model. This study used the probabilities and economic costs of Bond et al. (2007) and Bond et al. (2009a), and the gain in quality of life of 0.03 (Summerfield et al., 2003). This study concluded that the ICER of BCI lies below the WTP of NICE. However the sensitivity analysis showed that the probability of BCI being cost-effective was only 48%.

More recently Chen et al. (2014) analyzed the cost-effectiveness of BCI with respect non-implantation and with respect the UCI. The ICERs obtained were $14,658/QALY and $55,020/QALY respectively. In the first case BCI was clearly cost-effective, but in the second the conclusion was quite uncertain.

Table 6.1 shows a summary of the ICER obtained in these studies, which differ significantly. One factor that might explain the differences is the country where each study was conducted. Almost all direct and indirect costs of the treatment as well as its effectiveness vary according the geographic location.

Another factor that affects the ICER is the gain in quality of life used in each study. Table 6.2 shows the estimates of the gain in quality of life obtained by previous studies. This is the parameter with the highest influence on the results, and therefore can significantly change the obtained ICER.

In conclusion, the most recent studies tend to confirm that the BCI is cost-effective. However, the degree of uncertainty was still high when we decided to conduct our study of BCI in Spain.

[3]Range of ICERs obtained when applying different techniques for measuring incremental quality of life.

Study	Country	Sources	Population	ICER
Summerfield et al. (2002)	United Kingdom	own data	adults	£61,734/QALY
Summerfield et al. (2003)	United Kingdom	Summerfield et al. (2002)	adults	£100,000/QALY
Summerfield et al. (2006)	United Kingdom	own data	adults	£102,500/QALY
Barton et al. (2006)	United Kingdom	own data	adults	£32,909/QALY
			children	£39,049/QALY
Bond et al. (2007) Bond et al. (2009a)	United Kingdom	data projection	adults	£49,559/QALY
			children	£40,410/QALY
L-Pedraza Gómez et al. (2007)	Spain	Summerfield et al. (2002)	adults	€53,018/QALY
			children	€44,199/QALY
Bichey and Miyamoto (2008)	USA	own data	adults and children	$2,187/QALY
Summerfield et al. (2010)	United Kingdom	own data	children	£21,768/QALY
Chen et al. (2014)	Canada	own data	adults	$55,020/QALY
Kuthubutheen et al. (2015)	Canada	own data	adults	$(16,047 - 55,020)/QALY[3]

Table 6.1: Summary of previous cost-effectiveness studies of BCI.

Study	Population	Informants	Method	ΔQoL
Summerfield et al. (2002)	adults	experts	TTO	0,031
Summerfield et al. (2006)	adults	patients with BCI	HUI-3	0,03
Bichey and Miyamoto (2008)	adults and children	patients with BCI or their parents	HUI-3	0,11
Lovett (2010)	children	parents of children with BCI and UCI	VAS	0,33
Summerfield et al. (2010)	children	experts, students and parents of disabled children	VAS	0,06
			TTO	0,05
Sparreboom et al. (2012)	children	patients with BCI	HUI-3	0,04
Härkönen et al. (2015)	adults	patients with BCI	15D	0,03

Table 6.2: Summary of previous studies that estimate the gain in quality of life.

6.2 Model

We have built a state-transition model, implemented as an MID, which represents the events that may occur along the life of a CI user implanted at the age of one year. The cycle length is one year. We assume that the child is severely to profoundly deaf, has not been implanted previously, and receives one or two implants. We must take into account that since 2010 all the regional health systems in Spain apply screening programs to all newborns, and in case of severe to profound deafness, the child is implanted when he/she is around 12 months old.

Figure 6.3 shows the graph of the model. It contains one decision, *Intervention decided*, drawn as a blue rectangle, with three possible options: *simultaneous BCI* , *sequential BCI* and *UCI*. Given that in developed countries the standard practice is to give at least one implant to all children with severe to profound deafness, we did not include the option *no implant* in our model.

Chance nodes are drawn as rounded rectangles. The node *Intervention applied* represents whether the patient receives one implant, two implants simultaneously, or two implants sequentially. In general this variable takes the same value as *Implantation*

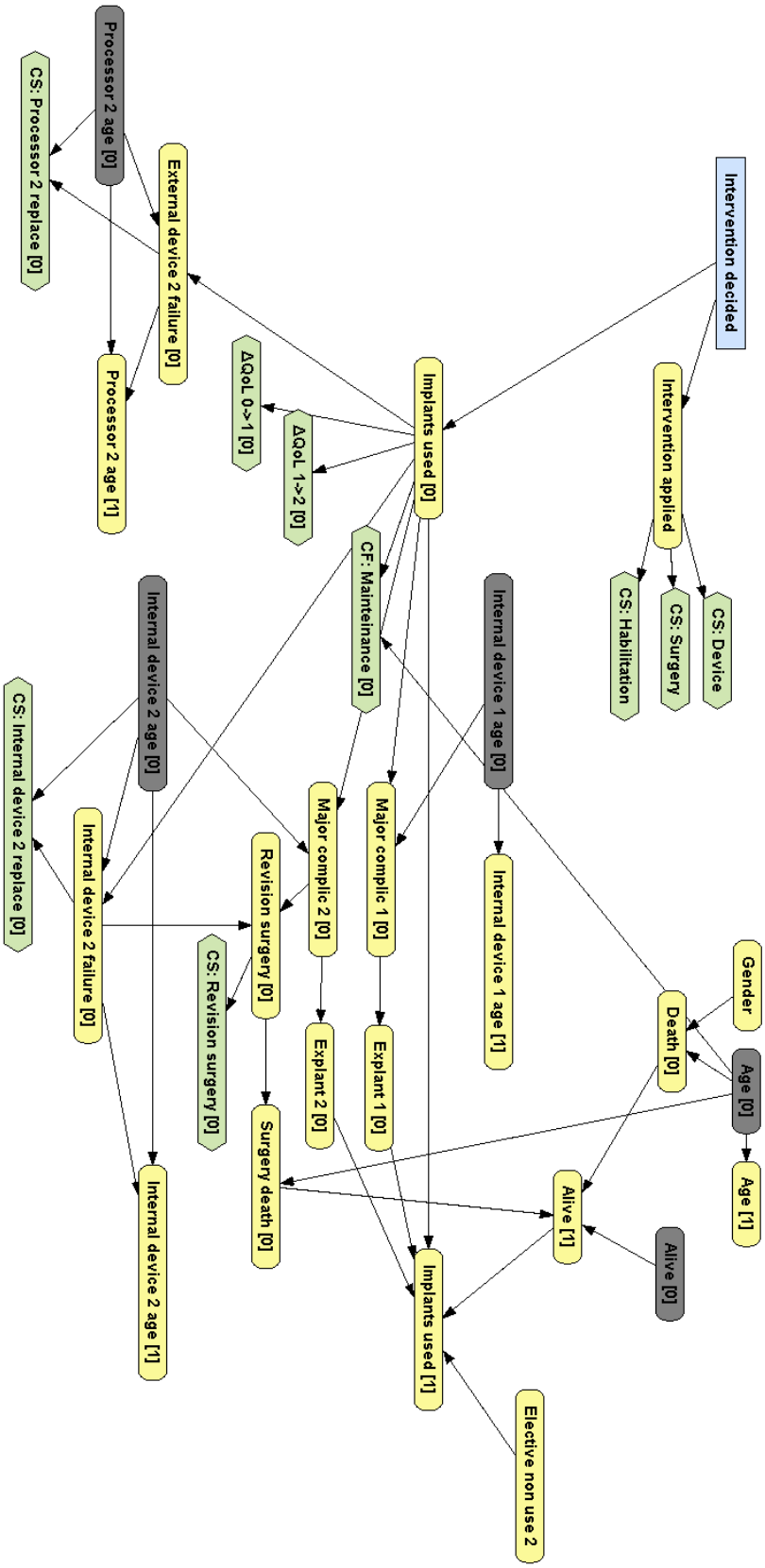


Figure 6.3: A Markov influence diagram for cochlear implantation.

decided, but when the decision is to put two implants simultaneously, sometimes they must be put in different operations due to clinical complications, as discussed below. The six nodes in the upper right zone of the graph in Figure 6.3 represent the gender and age of the patient, and whether he/she is alive. *Gender* does not have a temporal index because it is an atemporal variable, i.e., its value does not change. Variables whose value can change over time have indices, written in square brackets, which denote the period of time. Thus, *Implants used [i]* indicates how many implants the patient is using in the i-th cycle; it has four states: *both implants*, *only first implant*, *only second implant*, or *none*. The numbers 1 and 2 written in the name of a variable outside the brackets denote the first or the second implant respectively. For example, *Elective non use 2* indicates whether the person decides to use the second implant or not. When the value of a chance variable is known with certainty, the node is colored in gray; in this model we know with certainty that the child is one year old and alive (nodes *Age [0]* and *Alive [0]*), and all the external and internal components of the CI are brand new (nodes *Processor 2 age [0]*, *Internal device 1 age [0]*, and *Internal device 2 age [1]*).

Value nodes are drawn as green hexagons. The nodes $\Delta QoL_{0\to1}$ *[i]* and $\Delta QoL_{1\to2}$ *[i]* denote the increase in quality of life from no implant to one implant and from one implant to two implants respectively; the quality of life accrued in the i-th cycle depends on the number of implants used in it. The other value nodes represent monetary expenditures. The prefix "CS" means "cost for the (health) system" and "CF" means "cost for the family". The three child nodes of *Intervention applied* represent the costs of implantation; they are atemporal because they are paid only once. The other costs depend on the number of implants used in each cycle and on the occurrence of adverse events.

6.2.1 Probabilities

Every chance variable in the model has a conditional probability distribution for each configuration of its parents in the graph. The node *Gender* has no parent; therefore its probability is just the prior probability of being male or female. We obtained it from the database of the Spanish National Institute of Statistics (INE) for January 1st,

2015.[4] The probability of dying in the i-th cycle, represented by *Death [i]*, depends on *Gender* and *Age [i]*; it is therefore the annual mortality rate, also taken from the INE.

If *Intervention decided* takes the value *UCI* or *sequential BCI*, then *Intervention applied* takes the same value, with probability 1, but when *Intervention decided* = *simultaneous BCI*, it is not always possible to do it. Therefore, the probability of *Intervention applied* is 89% for *simultaneous BCI*, 10% for *sequential BCI*, and 1% for *UCI*. We assigned these probabilities because in the 50 cases reported by Ramsden et al. (2009) it was always possible to put both implants in one surgery, as planned, but in other 50 cases planned for simultaneous BCI (Holland et al., 2012), it was possible only in 39 patients; in 10 cases the second ear had to be implanted later and one patient received only one implant, for unknown reasons.

The probability tables for some nodes in the model are degenerate, i.e., the child variable depends deterministically on the values of its parents. For example, *Alive [i]* = *true* if and only if the user was alive in the previous cycle and has not died, i.e., if *Alive [i−1]* = *true* and *Death [i−1]* = *false*. Similarly, the value of *Implants used [0]* is determined by the intervention applied: if *Intervention applied* = *UCI*, then the patient is initially in the state *only first implant*, whereas if the intervention is either *simultaneous BCI* or *sequential BCI*, the initial state is *both implants*. In subsequent cycles, *Implants used [i]* takes the same value as in the previous cycle except when the user dies, when he/she decides not to use the second implant, or when the internal device has been explanted and could not be reimplanted. Even though there are adolescents and adults who refuse to use the second implant when they receive it a long time after the first one, the experts we consulted in Spain and at international conferences could not recall any case of a child who refused one or both implants after receiving them at a very early age; therefore those experts agreed that it is a good approximation to set that probability to zero in our model. We might have removed this node, but decided to maintain it for similarity with previous models (Bond et al., 2009b; Summerfield et al., 2010) and to perform a DSA for that probability.

The probability of a failure of the external components was set to 0.115 for the first two years, 0.095 between the third and the fifth, 0.104 for the sixth and seventh, and 0.16 afterwards, as in (Bond et al., 2009b).

[4] `www.ine.es/jaxiT3/Tabla.htm?t=9663`

Revision surgery may be due to a failure of an internal device or to a major medical/surgical complication. In accordance with (Bond et al., 2009b; Summerfield et al., 2010) we set the probability of internal device failure to 0.001 for the first 9 years and 0.003 afterwards, and the probability of major complication (other than a failure) to 0.041 in the first year and 0.004 afterwards. Some manufacturers have reported lower failures rates,[5] but due to discrepancies in their reports and in published studies (Rădulescu et al., 2013; Hildrew and Molony, 2013), we decided to use the same probabilities as in (Bond et al., 2009a; Summerfield et al., 2010), thus taking a conservative position in favor of UCI[6]. We also followed those authors in assuming that a failing internal device can be replaced, but a major complication may make it impossible to reimplant the device; we set the probability of permanent explantation in case of major complication to 0.043, in accordance with (Wang et al., 2014).

The possibility of dying at revision surgery, mainly because of the complication of anesthesia, slightly reduces the effectiveness of bilateral implantation. The perioperative mortality rate is less than 1/10,000 for children (González et al., 2012) and less than 20/10,000 for adults (Braz et al., 2009). Taking a conservative approach in favor of UCI, we set those probabilities to 0.0001 until the age of 18 and 0.002 afterwards.

Table 6.3 summarizes the probabilities used in the model.

6.2.2 Outcomes and costs

The effectiveness depends on the quality of life at each moment, $QoL(t)$ (Equation 3.1), which in our model only depends on the number of implants used, as shown in Figure 6.3. Therefore, the incremental effectiveness of BCI with respect to UCI mainly depends on the difference in quality of life between having one and two implants, $\Delta \text{QoL}_{1\to 2}$, as we explained in Section 6.1.3. In our general-population survey, the estimates for this parameter ranged from 0.106 to 0.293, depending on the elicitation technique (Artaso

[5] `www.medel.com/int/reliability-reporting`, `www.advancedbionics.com/content/dam/ab/Global/en_ce/documents/candidate/AB_Reliability_Report.pdf`

[6] Given that our final goal was to convince the Spanish health authorities that BCI is cost-effective, we made several conservative modeling decisions in favor of UCI, to make the conclusions more compelling.

Parameter	Value	Source
Being male	0.4914	National Institute of Statistics
Major complication | device age $<$ 1 yr	0.041	Bond et al. (2007) Summerfield et al. (2010)
Major complication | device age $\geq$ 1 yr	0.004	Bond et al. (2007) Summerfield et al. (2010)
Perm. explantation | major complication.	0.043	Wang et al. (2014)
Surgical mortality | age $<$ 18 yr	1.0E-4	González et al. (2012)
Surgical mortality | age $\geq$ 18 yr	0.002	Braz et al. (2009)
External device failure | 1^{st}yr	0.16	Bond et al. (2007)
External device failure | 2^{nd}yr	0.115	Bond et al. (2007)
External device failure | 3^{rd}- 4^{th} yr	0.095	Bond et al. (2007)
External device failure | 5^{th}- 6^{th} yr	0.104	Bond et al. (2007)
External device failure | $\geq 7^{\text{th}}$yr	0.16	Bond et al. (2007)
Internal device failure | $< 8^{\text{th}}$yr	0.001	Bond et al. (2007)
Internal device failure | $\geq 8^{\text{th}}$yr	0.0035	Bond et al. (2007)

Table 6.3: Probabilities used in the model. The annual mortality rate, which is not shown in this table because it depends on age and gender, was also taken from the National Institute of Statistics.

and Díez, 2016). In the model we used the lowest value, $\Delta\text{QoL}_{1\to2} = 0.106$, for two reasons: because that was the value obtained with the version of the time trade-off method most commonly used in quality of life studies, and because we wished to take a conservative approach in favor of UCI. As in our survey we did not measure the increase in quality of life from no implant to one implant, $\Delta\text{QoL}_{0\to1}$, we took the value 0.106 from (Summerfield et al., 2010).

According with Equation 3.3, the ICER does not depend on the absolute costs of the three interventions—simultaneous BCI, sequential BCI, or UCI—but on the differences between them. For this reason we have not explicitly included in our model the costs common to the three interventions, such as the cost of the first implant. This way we avoid including in the model parameters which are absolutely irrelevant.

The price of the second implant was set to €21,000, the average of the prices we obtained from a manufacturer and a distributor for our country[7]. We did not assume any discount, for the reasons discussed below.

In the case of simultaneous BCI, we assumed that surgery requires 100 more minutes than UCI (Smulders et al., 2016) at a cost of €229/hour, including personnel and materials; this value is the average of €169/h (Herranz Amo et al., 2006) and €289/h (Palà et al., 2003). Sequential BCI has an extra cost of €1,000 for the pre-surgical selection and exploration assessment (50% of the corresponding cost for UCI), plus €3,500 for the surgical intervention and €600 for rehabilitation (Torre et al., 2005), the same as for UCI. These are conservative estimates in favor of UCI, because in general the pre-surgical assessment made for UCI is also valid for the second implantation, and because if the gap between implantations is short, both ears can be rehabilitated at the same time.

If the internal device fails during the first ten years, the manufacturer provides a new one; after the guarantee period, the health system pays €13,000 for it (communication from a manufacturer). In all cases the health system takes over the cost of surgery.

If the external processor fails during the first two years, it is replaced by the manufacturer. After the seventh year, it is replaced by the health system.[8] Between

[7]The other main manufacturer did not respond when we asked them.

[8]Even though a national law says that a processor that fails after seven years of use must be replaced,

the third and the sixth year, the user must cover the repairs; we have included this expense in maintenance costs, together with coils, microphones, cables, and batteries, which in Spain are covered by users and their families. According to our study with CI users, these costs amount to €675/year for users under the age of 18 and €373/year for adults.

In our model we did not include the incremental cost of programming the second implant because in most cases it is covered by manufacturers, whose experts travel to hospitals in different cities; in the few cases in which the processors are programmed by the personnel of the implantation center, the marginal cost is low. We will come back to this issue in the discussion.

Given that all the cost estimates were obtained in few months before submitting our study to a journal, we did not apply any conversion to current euros.

Table 6.4 summarizes the quality of life increases and the economic costs used in our model.

6.2.3 Second-order uncertainty

The uncertainty of the model was represented by adding second-order distributions to the main parameters (Briggs et al., 2006). Each parameter denoting a probability p (in the reference case) was assigned a beta distribution $B(\alpha, \beta)$ with a mean of p, i.e., $\alpha/(\alpha + \beta) = p$; when p was taken from a study that did not indicate the number of subjects involved, we assumed a sample size of 100, which is implemented by making $\alpha + \beta = 100$, as in (Luque et al., 2016). Therefore $\alpha = 100p$ and $\beta = 100(1 - p)$.

We modeled the increment in QoL from one to two implants, $\Delta\text{QoL}_{1\to2}$ with a Gamma with $k = 22.832$ and $\theta = 0.004$, as this is the distribution that better fitted the data from our general-population survey (Artaso and Díez, 2016). The increment from total deafness to one implant, $\Delta\text{QoL}_{1\to2}$, was modeled with a Gaussian, $\mathcal{N}(0.1056; 0.1217)$, in accordance with (Summerfield et al., 2010).

Following the recommendation of Briggs et al. (2006), costs were modeled with

the replacement period is often longer because some regional governments, especially in these times of economic crisis, are reluctant to fulfill their duties. However, in our model we have assumed that all of them comply with the regulation.

Parameter	Value	Source
$\Delta QoL_{0\to1}$ (from zero to one)	0.106	Summerfield et al. (2010)
$\Delta QoL_{1\to2}$ (from one to both)	0.106	Artaso and Díez 2016
CS: Device	€21,000	manufacturers
CS: Surgery (increase for simultaneous BCI)	€382	Palà et al. (2003) Herranz Amo et al. (2006) Smulders et al. (2016)
CS: Surgery (increase for sequential BCI)	€4,500	Torre et al. (2005)
CS: Habilitation (increase for sequential BCI)	€600	Torre et al. (2005)
CS: Processor 2 replace	€6,300	manufacturers
CS: Internal device 2 replace	€13,000	manufacturers
CF: Maintenance (each implant)	€675 (age $<$ 18) €373 (age $\geq$ 18)	our survey among CI users

Table 6.4: Quality of life and economic costs.

Gamma distributions. We used a Gamma with $k = 155.961$ and $\theta = 4.314$ for the maintenance costs of users under 18, and $k = 37.491$ and $\theta = 9.866$ for adults, following the distributions that better fitted the data collected by our survey (Artaso and Díez, 2016). In the remaining cases, each parameter denoting a cost c was assigned a Gamma distribution with a mean $\mu = c$ and a standard deviation $\sigma = \mu/5 = 0.2c$, as in (Bond et al., 2007; Summerfield et al., 2010).

These choices of equivalent sample size and standard deviations are debatable, but the sensitivity analyses showed that their impact is virtually null for a WTP of €30,000/QALY.

6.2.4 Evaluation of the model

We expanded the MID for a horizon of 100 cycles, i.e., until the age of 101, and evaluated the model with the algorithms, described in Section 5.2. Following the recommendations of the *Panel on Cost-Effectiveness in Health and Medicine* convened by the U.S. Public Health Service (Gold et al., 1996), we applied an annual discount rate of 3% on both costs and effectiveness—the same value as in other economic evaluations of BCI (L-Pedraza Gómez et al., 2007; Chen et al., 2014; Catalá-López et al., 2016)—and in the DSAs we applied rates of 0%, 5% and 7%.

We also performed a PSA with 10,000 simulations in order to obtain a scatter plot and an acceptability curve (Drummond et al., 2005; Briggs et al., 2006), and other types of DSAs to find out how the ICER varies for different parameters of the model.

6.3 Results

6.3.1 Reference case analysis

When considering only the costs currently covered by the Spanish health system, the incremental effectiveness of simultaneous BCI with respect to UCI was 3.163 QALYs, and the incremental cost €32,649, which leads to an ICER of €10,323/QALY. For sequential BCI the effectiveness is the same as for simultaneous BCI, the incremental cost is €37,109, and the ICER €11,733/QALY. If the health system covered all the maintenance costs (at the current prices), as in other countries, the ICER with respect to UCI would be €15,035/QALY for simultaneous BCI and €16,446/QALY for sequential BCI.

These values are far below the WTP of the Spanish health system, that is estimated on €30,000/QALY as discussed in Section 3.1.2, which implies that BCI is clearly cost-effective in all cases.

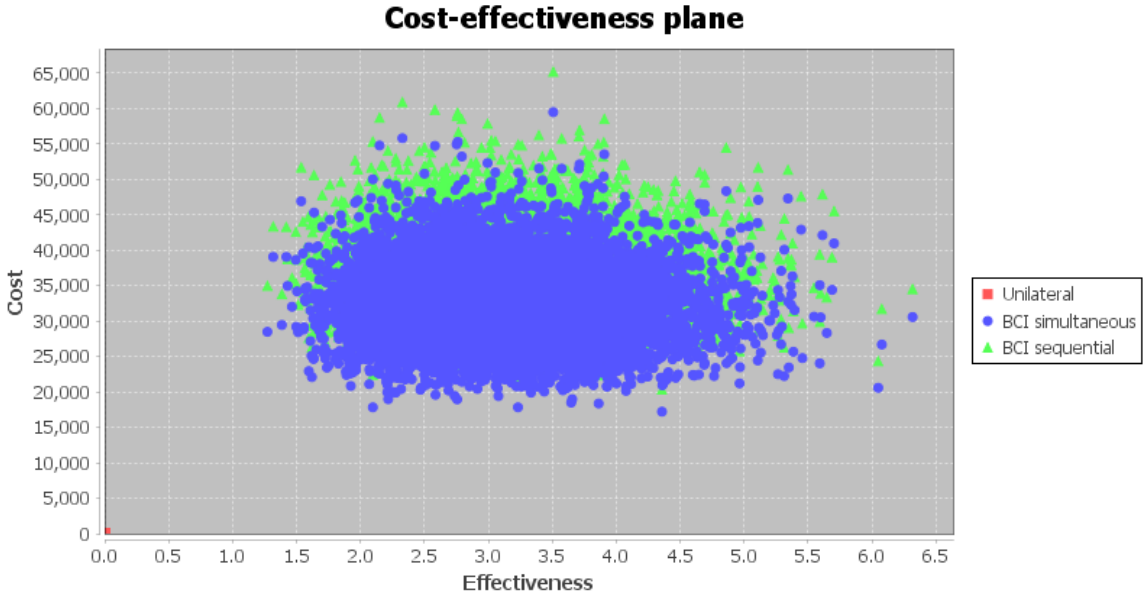


Figure 6.4: Cost-effectiveness scatter plot for simultaneous and sequential BCI, including only the costs currently covered by the Spanish health system.

6.3.2 Sensitivity analyses

Figure 6.4 shows the scatter plot resulting from the PSA, with UCI as a comparator. As a consequence of the hypotheses in our model, sequential BCI always has the same effectiveness as simultaneous BCI, but at a higher cost. In all the simulations the ICER was below €30,000 with respect to UCI, which implies that for this WTP both simultaneous and sequential BCI are cost-effective with virtually 100% certainty. This conclusion can also be drawn by observing the acceptability curves in Figure 6.5.

If the health system covered all the maintenance costs, the acceptability curves for BCI would increase more slowly (see Fig. 6.6), but it is almost sure that both interventions are cost-effective with respect to UCI—the probability is 99.98% for simultaneous BCI and 99.92% for sequential BCI.

Given that the most influential and most uncertain parameter in the model is the increase in quality of life from UCI to BCI, represented by the node $\Delta QoL_{1\to 2}$ in Figure 6.3, we performed a DSA for it, assuming that the other parameters take the same values as in the reference case. We can observe in Figure 6.7 that for a WTP of €30,000/QALY, an increase of QoL as low as 0.037 suffices to make simultaneous

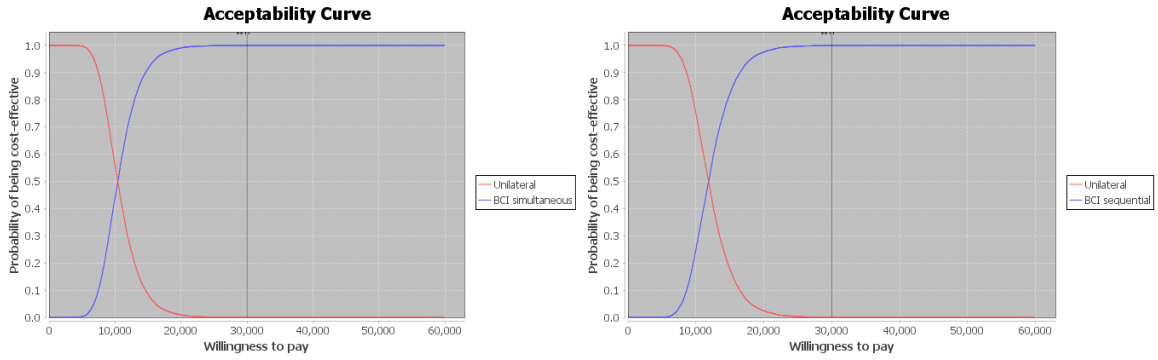


Figure 6.5: Acceptability curves for simultaneous (left) and sequential BCI (right), including only the costs currently covered by the Spanish health system.

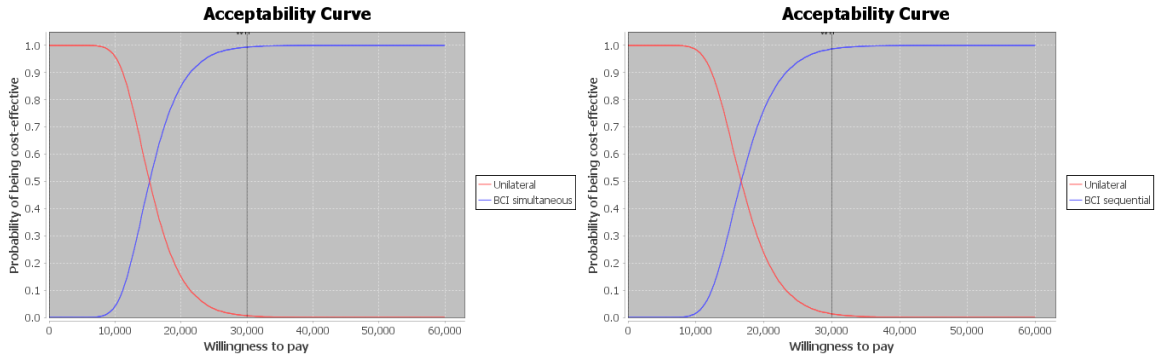


Figure 6.6: Acceptability curves for simultaneous (left) and sequential BCI (right) if the health system covered all the maintenance costs.

BCI cost-effective[9]; if the health system covered all the maintenance costs, it would be cost-effective for an increase of 0.054.

There is also uncertainty about the quality of life increase from no implant to one implant, $\Delta\text{QoL}_{0\to1}$, but this parameter has little influence on the results. Its reference value is 0.106. If we vary it from 0.05 to 0.20 (while maintaining the value of $\Delta\text{QoL}_{1\to2}$), the ICER only varies from 10,354 to 10,269 euros per QALY gained.

Another assumption in our model is that the probability of voluntary non-use of the second implant is null for children that receive both implants at the age of one. As mentioned above, the experts we consulted considered this a good approximation. Additionally, if we increased that probability to a value as high as 1%, the ICER of simultaneous BCI would only increase by €69/QALY.

[9]Let us remember that the value of ΔQoL obtained in our study (Artaso and Díez, 2016) is 0.106

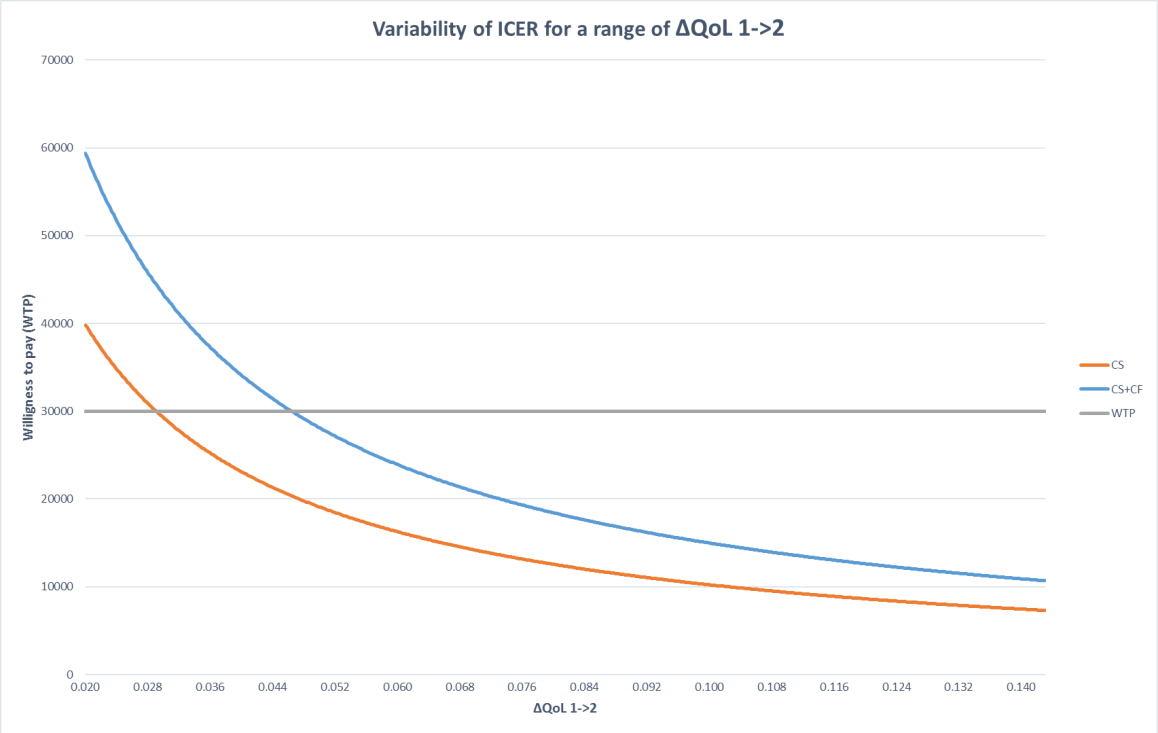


Figure 6.7: Sensitivity analysis: ICER of simultaneous BCI with respect to UCI as a function of the increase in quality of life from having one implant to having two, $\Delta\text{QoL}_{1\to 2}$. Under the current coverage policy (orange curve), BCI is cost-effective when $\Delta\text{QoL}_{1\to 2} > 0.037$. If the system covered all the maintenance costs (blue line), BCI would be cost-effective for $\Delta\text{QoL}_{1\to 2} > 0.054$.

We have also analyzed the impact of different discount rates. They are more relevant for CI than for other health technologies because a significant part (40%) of the cost is paid at the moment of implantation while the benefit (the effectiveness) is gathered along the user's lifetime, up to a horizon of 100 years. Consequently, if we raise the discount from 3% to 5% and 7%, as recommended in (Gold et al., 1996), the ICER for simultaneous BCI increases from 10,323 to 13,703 and 17,353 euros per QALY respectively, because discounts affect more the effectiveness than the costs. In the absence of discounts, the ICER would be €6,422/QALY. In all cases the ICER is clearly below €30,000/QALY.

6.4 Discussion

Our study has focused on one-year-old children with bilateral severe to profound hearing loss who qualify for bilateral CI. Unlike other studies (Bond et al., 2009b; Summerfield et al., 2010; Foteff et al., 2016b), the only comparator in our analysis was UCI; we did not compare BCI with "UCI plus a contralateral hearing aid" because the benefit of a hearing aid depends on the degree of hearing loss in that ear; in some cases a hearing aid provides better audition than a CI, while in others it is almost useless.

Simultaneous BCI clearly dominates sequential BCI because it is as safe and effective, but less expensive, even though we have not included in the model the extra cost assumed by users and their families (traveling, work hours lost, etc.). Therefore children that need two implants and have none should receive them in one surgical operation. However, because of current policies, many children have received only one implant. If they receive the second within a short time interval—say less than a year—it will be cost-effective according to our model, but the certainty of this conclusion depends on two parameters: the increase in quality of life from one to two implants and the WTP. Additionally, the effectiveness of the second implant decreases when the age of implantation and the time lag increase. Unfortunately, the lack of detailed effectiveness data in the literature makes it impossible to specify with precision in which cases sequential BCI is cost-effective. The dependence of cost-effectiveness with age of implantation has only been studied for UCI (Semenov et al., 2013).

Our results agree with those of recent studies (Bond et al., 2009b; Summerfield et al., 2010; Foteff et al., 2016b) and with the guidelines of the British National Institute for Health and Care Excellence, which recommends simultaneous BCI for newborn children and admits the possibility of a second implant for children who already have one "only if this is considered to provide sufficient benefit by the responsible clinician after an informed discussion with the individual person and their carers" (NICE, 2009). The main difference is that in our study the probability of BCI being cost-effective is much higher than in previous PSAs (Bond et al., 2009b; Summerfield et al., 2010; Foteff et al., 2016b).

The principal limitation of our analysis—and of all other CEA of BCI—is the uncertainty about the increase in quality of life from one to two implants, $\Delta\text{QoL}_{1\to2}$

(see Table 6.2). The studies aimed at measuring it have returned very disparate estimates, partly due to the use of different elicitation techniques. The values obtained using Visual Analog Scales (VAS) range from 0.02 (Smulders et al., 2016) to 0.33 (unpublished study by R. Lovett, cited in (Summerfield et al., 2010)); however, VAS estimates are not appropriate for economic evaluation because they are not based on a preference scale. The estimates obtained using quality of life indexes range from -0.003 (Lovett, 2010) to 0.11 (Bichey and Miyamoto, 2008); these extreme values are both obtained using the Health Utility Index, HUI-3. A problem of general health indexes is their low sensitivity to improvements in audition (Lutman, 2008; Sparreboom et al., 2012; Kuthubutheen et al., 2015; Smulders et al., 2016). In our opinion, the most reliable technique of those used for measuring that increase is the Time Trade-Off (TTO), because it is preference-based and does not suffer from a lack of sensitivity. Previous estimates of $\Delta\text{QoL}_{1\to 2}$ with the TTO have returned the values 0.031 (Summerfield et al., 2002), 0.05 (Summerfield et al., 2010), 0.12 (Kuthubutheen et al., 2015), and 0.09 (Smulders et al., 2016). In the survey we conducted for this project, the values obtained with the TTO ranged from 0.106 to 0.293 , depending on the way the questions were framed (Artaso and Díez, 2016); as mentioned above, in this research we have used the lowest of these values to make the conclusions more credible. Given that all the five TTO studies have returned values above the threshold of 0.042 derived in our sensitivity analysis (cf. Fig. 6.7), we can conclude that sequential BCI is cost-effective for the Spanish public health system beyond any reasonable doubt.

The disparity of TTO results makes it inappropriate to conduct a meta-analysis, but in order to have a rough summary, we note that the average is 0.078 and the median 0.09. As these values satisfy the condition $\Delta\text{QoL}_{1\to 2} > 0.054$ (see again Fig. 6.7), it is also reasonable to conclude, albeit with a small amount of uncertainty, that BCI would still be cost-effective if the health system decided to cover all the maintenance costs.

There are minor sources of uncertainty that we have not considered in our sensitivity analyses. One of them is the risk of meningitis as a consequence of CI. Some years ago that was a serious concern, but nowadays its incidence has decreased (Lalwani and Cohen, 2011; Chen et al., 2013), mainly because many cases were due to electrode positioners that are no longer used. Additionally, some of those cases are not due to the implant itself, but to "congenital abnormalities of the cochlea which

predispose [deaf people] to meningitis even prior to implantation" (FDA, 2002). In Spain all children are vaccinated against several types of meningitis, and some receive prophylactic perioperative antibiotic treatment before CI surgery, with no extra cost in the case of simultaneous BCI and a negligible cost for sequential BCI. For these reasons we assumed that BCI neither increases significantly the risk of meningitis nor requires additional expenditures to prevent it.

With respect to costs, some of the assumptions in our model are specific for current policies in Spain: in most implantation centers the cost of programming is covered by manufacturers; users and their families assume most of the maintenance costs, including batteries and cables, lost coils, the repairs of the processor (and its replacement, if necessary) between the third and the sixth year of use; the cost of surgery when the internal device is always covered by the health system, even during the guarantee period, etc. Additionally, in our model the second implant has the same price as the first one. In contrast, the technical report of Bond et al. (2007, Fig. 26) took into account the discounts on the second device, which in the UK might be "equivalent to 40% or more for the second implant" (NICE, 2009) and were decisive for the approval of BCI in that country. In this aspect our model agrees with most CEA of BCI, including the journal version of the paper by Bond et al. (2009b), which did not consider the effect of discounts because "the continued presence and size of these discounts in the future is impossible to guarantee".

6.5 Conclusions

Our study has proved beyond any reasonable doubt that BCI is cost-effective for the Spanish public health system. Even if the system took over all the maintenance costs, which in Spain are partially covered by users and their families, it would still be cost-effective, but with a small amount of uncertainty, mainly due to the imprecise estimation of quality of life increase from one to two implants. From the societal perspective, BCI would be even more cost-effective because of savings in education and better employment opportunities, but there is not enough data to quantify these advantages.

Given that putting both implants in one surgical operation saves costs for

the health system (additional surgery and (re)habilitation) and for users and their families (traveling, work hours lost...), bilateral implantation should be performed simultaneously. For little children who already have one implant, it is cost-effective to give them a second one. In this case, the benefit decreases with age and with the gap between implantations, but again the uncertainty about the increase in quality of life makes it impossible to specify with precision when sequential BCI ceases to be cost-effective.

Our analysis agrees with recent studies that concluded that BCI is cost-effective for children (Bond et al., 2009b; Summerfield et al., 2010; Foteff et al., 2016b), at least when performed simultaneously, and possibly also for adults (Chen et al., 2014; Kuthubutheen et al., 2015; Foteff et al., 2016a; Smulders et al., 2016), but disagrees with previous studies carried out in Spain, which asserted, after brief analyses, that it was not cost-effective even for children (Estrada et al., 2011; L-Pedraza Gómez et al., 2007).

The conclusions of this study are specific for Spain. We can assume that health preferences, and consequently the effectiveness of BCI, are very similar in other countries, but costs may differ significantly because of variations in both prices and policies. In (Pérez-Martín et al., 2017a) we announced that the MID used for this study is publicly available at `www.probmodelxml.org/networks`, so that other researchers—using our open-source software tool OpenMarkov—can examine the model, replicate the results, perform other sensitivity analyses, and adapt it to different countries.

7

Cost-effectiveness of colorectal cancer screening

"Medicine is a science of uncertainty and an art of probability."

William Osler

The work described in this chapter was carried out in collaboration with Marta Lalana of the Clinical Analysis Service of the Hospital de Barbastro in Huesca, Spain. The implementation of the CRC detection programs in Aragon and all the data collected in that study were carried out before this work. We will refer to that work as the "Aragon study". Partial results of this work were presented in (Lalana et al., 2016b; Lalana et al., 2016a).

7.1 Introduction

7.1.1 Colorectal cancer

CRC is one of the most frequent causes of morbidity and mortality in the world. In 2012 more than 1.3 million new cases were diagnosed and near 700,000 people died for this cause (International Agency for Research on Cancer, 2012). In Spain, CRC is the most frequent cancer with 41,441 new cases and a 15,449 deaths in 2015 (Sociedad Española de Oncología Médica, 2017).

CRC can be localized in the colon, the rectum, or both (see Fig. 7.1). In most cases, CRC starts as an *adenomatous polyp*, a small set of noncancerous cells, which may turn into CRC. Adenomas can be small and cause few or no symptoms, but finding and removing them can prevent CRC.

There are two main staging systems used to describe the state of CRC. In the Tumor-Node-Metastases (TNM) system, developed by the AJCC (American Joint Committee on Cancer), the stage is determined by three variables: tumor “T”, regional nodes “N”, and metastases “M” (Edge and Compton, 2010). The second system classifies the state of CRC into 5 stages, from 0, an early stage of CRC, to IV, in which cancer has spread to distant parts of the body.

7.1.2 Screening programs

Screening consists in looking for a disease in asymptomatic people. It may find the disease at an early stage or even prevent it, for example CRC can be avoided by

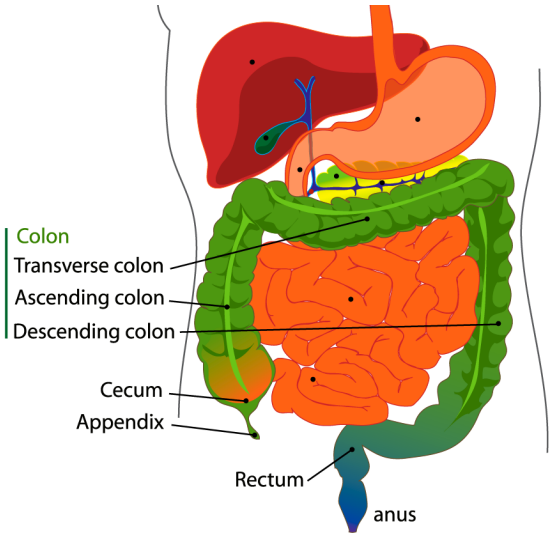


Figure 7.1: Graphical representation of the colon and rectum.

removing the polyps before they become tumors. Usually an abnormal result in the screening is not decisive but may lead to performing a diagnostic test, i.e. a test that can diagnose the disease with relative certainty. General population CRC screening programs exclude people having high-risk factors, such as previous CRC, family history of CRC, Lynch Syndrome, etc. Those people are assigned to more intensive surveillance diagnostic testing as colonoscopy.

Not all the diseases are suitable for screening programs. CRC meets the necessary requirements for being subjected to screening: known natural history, high incidence, morbidity and mortality, and availability of a test, the Fecal Occult Blood Test (FOBT), which is simple, reliable, and efficient. Randomized Control Trials (RCTs) have proved that screening with FOBT reduces the mortality between 15 and 33% (Kronborg et al., 1996; Hardcastle et al., 1996; Towler et al., 1998; Faivre et al., 2004). CRC screening with FOBT improves the prognosis, with the consequent reduction of incidence and mortality (Gómez Hernández et al., 2013).

The introduction of a CRC screening program implies inevitably an immediate increase in costs, but also to mid and long-term saving due the decrease in the incidence of CRC and the avoidance of expensive treatments.

In Spain, the updating of the cancer strategy of the Spanish National Health System includes among its objectives the implementation of population-based CRC

screening programs for individuals with low to medium risk.[1] Since 2016 all the autonomous communities, except Ceuta and Melilla, have implemented a program consisting in measuring every two years the concentration of hemoglobin occult in feces and performing the colonoscopy when the test is positive. In Aragon, the program started in 2014 for a target population between 50 and 69 years old. In 2015 the coverage reached 19% of this population (see fn. 1).

The goal of the study described in this chapter was to determine the cost-effectiveness of CRC screening program with a biennial FOBT. We consider two possible ages of entry, 50 and 60 years, and compared them with the no-screening intervention.

7.2 Model

The model was implemented as an MID (Díez et al., 2017) using OpenMarkov. Figure 7.2 shows the compact representation of the model.

It is a state-transition model, based on the natural history of CRC (Tappenden et al., 2007; Sobhani et al., 2011), which evolves from normal epithelium to malignant lesion. The five health states: normal epithelium, low-risk adenoma, high-risk adenoma, CRC, and dead, are modeled with the variable *State [i]*. The stage of the CRC is modeled with another variable, *CRC stage [i]*, as shown in Figure 7.3.

Our model assumes that the cohort begins with "normal epithelium" at the age of 30 years and moves through the states of the model, with a cycle length of one year and a temporal horizon of 70 cycles, that is, until the age of 100 years.

Following the classification of the American Cancer Society, polyps are considered "low-risk adenomas" when they meet the following conditions: presence of one or two adenomas less than 10 mm, tubular, and low grade. In our model, we unify high and medium-risk adenomas into a single health state, which we call "high-risk adenomas". They are those that meet any of the following conditions: presence of three or more adenomas smaller than 10 mm; one or more adenomas with a size equal to or greater

[1] `http://www.cribadocancer.com/index.php/cancer-colorrectal/red-de-programas-de-cribado-espanoles/situacion`

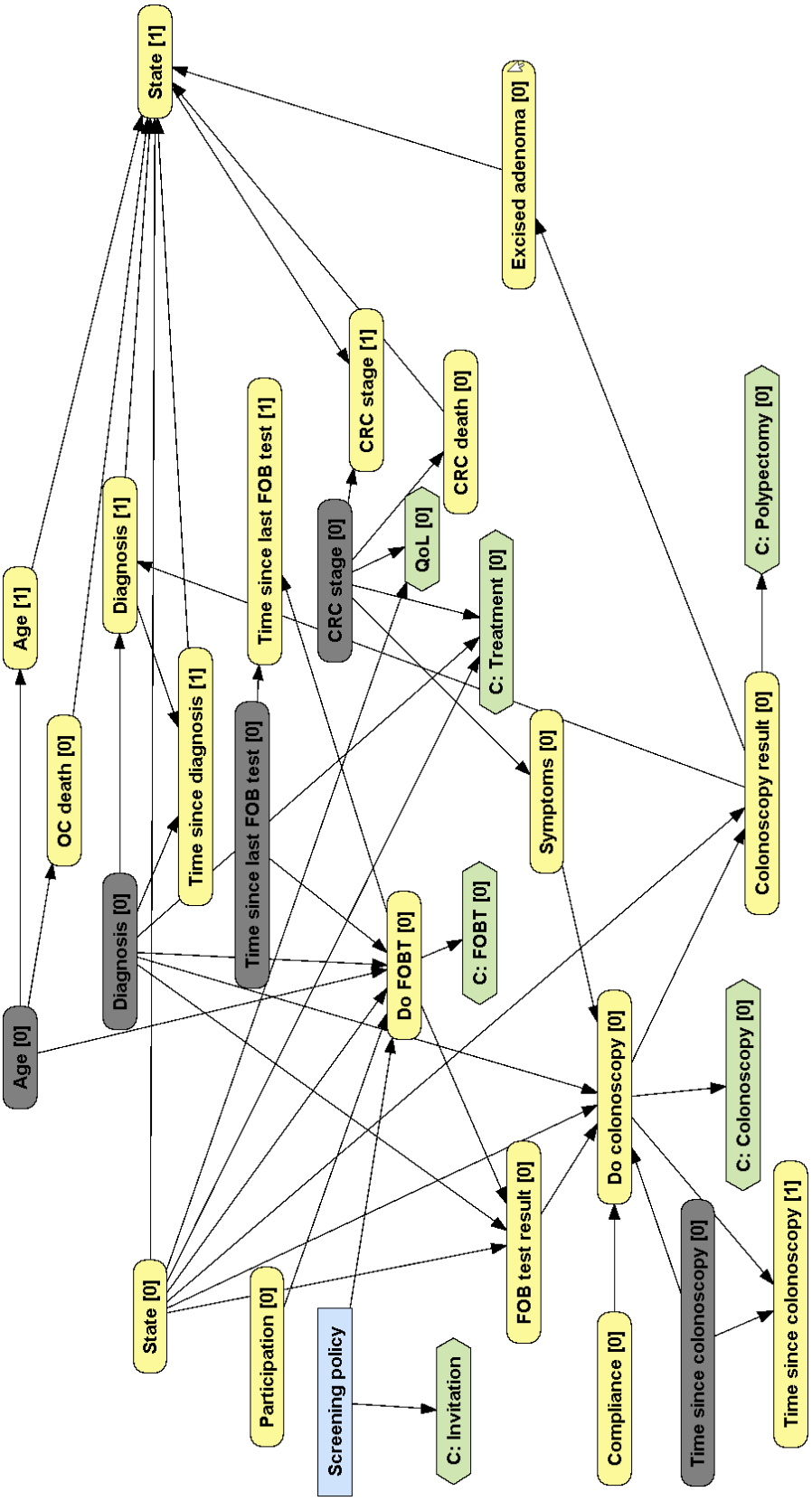


Figure 7.2: MID for CRC screening.

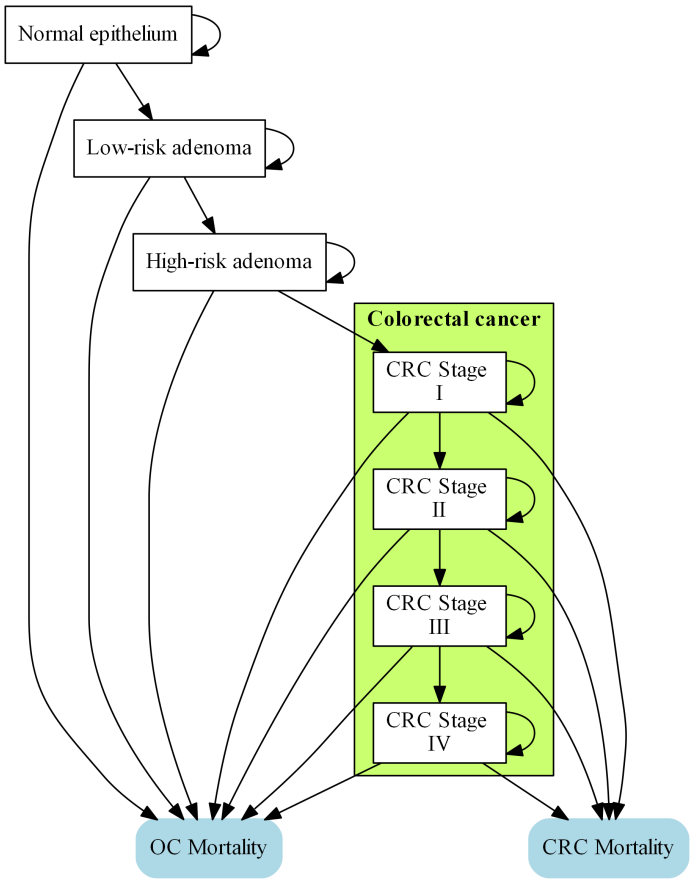


Figure 7.3: State-transition model based on the natural history of CRC.

than 10 mm, or with high-grade mucosal dysplasia, or with 20% or more of fluffy component. This group also includes carcinomas in situ (intraepithelial carcinoma or that invades the lamina propria).

Adenomas do not present spontaneous regression. Those less than 1 cm in diameter can grow and become CRC. We have considered that CRC can only be produced from a high-risk adenoma.

When FOBT is positive, the patient must be recommended for colonoscopy, as shown in Figure 7.4. The adenomas detected are removed at the moment with a polypectomy. If they turn out to be high risk, a colonoscopy is performed three years later; if they are low risk is performed five years later. Patients with a negative colonoscopy are taken out of the screening program, and a follow-up colonoscopy is performed 10 years later, as recommended by the CRC screening guidelines (Levin

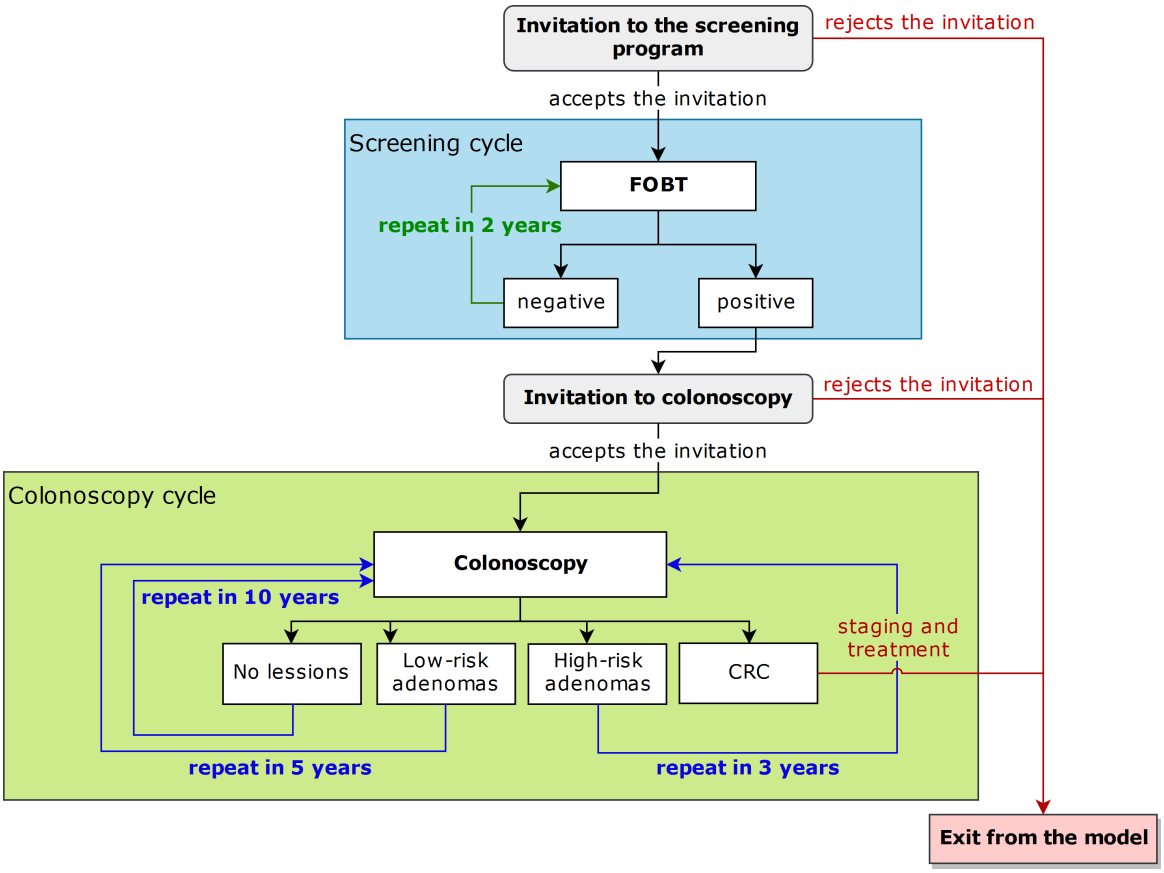


Figure 7.4: Scheme of the CRC screening program.

et al., 2008).

Patients diagnosed with CRC, either by screening or by the appearance of symptoms—even before the start of screening—are treated and then classified into clinical follow-up stages according to their TNM.

7.2.1 Probabilities

The annual transition probabilities for our model were obtained from the literature. From normal epithelium to low-risk adenoma the probability is 1.60%; from low-risk adenoma to high risk adenoma, 2.12%; from high-risk adenoma to CRC Stage I, 3.26%; from Stage I to Stage II, 58.29%; from Stage II to Stage III, 65.55% and from Stage III to Stage IV, 86.48% (Tappenden et al., 2007).

Parameter	Value	Source
Sensitivity of FOBT to adenomas	28.2%	Gondal et al. (2003)
Sensitivity of FOBT to CRC	68.8%	Allison et al. (1996)
Specificity of FOBT	94.4%	Allison et al. (1996)
Sensitivity of colonoscopy to low-risk adenomas	86.7%	Martín-López et al. (2014)
Sensitivity of colonoscopy to high-risk adenomas	92.9%	Martín-López et al. (2014)
Sensitivity of colonoscopy to CRC	94.7%	Martín-López et al. (2014)
Specificity of colonoscopy	100%	López-Bastida et al. (2010) Martín-López et al. (2014)

Table 7.1: Sensitivity and specificity of FOBT and colonoscopy.

The recurrence rate of low-risk adenoma—i.e. the probability of developing a new adenoma after the polypectomy—is 18% in the first year and 5% in subsequent years. After the removal of a high-risk adenoma this probability is 25% for the first year and 6% for the subsequent years (Winawer et al., 1993).

The sensitivity and specificity data of the FOBT and the colonoscopy are shown in Table 7.1, indicating the bibliographic source of each probability.

The probability of clinical symptoms is 7% for CRC Stage I, 32% for Stage II, 49% for Stage III and 85.4% for Stage IV (Tappenden et al., 2007).

The annual specific mortality, represented by the node *CRC_death [i]*, is: 0% for Stage I, 1% for Stage II, 6.02% for Stage III and 38.37% for Stage IV (Tappenden et al., 2007). The probability of death from other causes, node *OC_death [i]* in the model, was obtained from the Spanish National Institute of Statistics (INE),[2] subtracting from the general mortality rate the cases of CRC.

We have estimated the participation and compliance rates through the "Aragon study", registered patients from 8 health centers which from January 2014 to June 2015: 3 centers from the Barbastro sector (Abiego, Berbegal and Barbastro) and 5 from Zaragoza III (Delicias Norte, Miralbueno, Cariñena, Borja and Bombarda). They comprise a total of 83,699 inhabitants (Amorín Calzada, 2008),[3] of whom 12,122 were between 60 and 69 years old. In a first filtering, 1,428 persons were excluded due to

[2] `www.ine.es/dynt3/inebase/index.htm?type=pcaxis&file=pcaxis&path=/t20/e301/defun/a2010`

[3] `http://zaragoza3.es/Gerencia/Poblacion/pob_sector.htm`

clinical criteria by using data stored in health centers databases; such as previous digestive pathologies, family history of CCR, etc... The rest, 10,694 persons, were invited to participate in the study. After the interview conducted in the nursing consultation, 588 persons were excluded for other clinical reasons and 135 refused to participate in the study. Therefore, from the 9,971 persons eligible for screening, 5,874 participated, representing a rate of 58.91%. Only one patient with a positive FOBT refused colonoscopy, which means a compliance of 99.82%.

7.2.2 Outcomes and costs

Because of the scarcity of data about quality of life related to CRC in our country, we used the estimates obtained by Ness et al. (1999) in the United States: 0.91 for the general population; 0.74 for CRC Stage I; 0.70 for Stage II; 0.50 for Stage III, and 0.25 for Stage IV.

Our model includes two types of costs: those of the screening program and those associated with the diagnosis, treatment, and monitoring of CRC.

The cost of the invitation to the program, €0.65, was obtained from (Sobhani et al., 2011). The cost of FOBT, including the transfer of the analyzer, reagents, and maintenance, is €1.67, according to the framework agreement signed by the Health Service of Aragon and Sysmex Spain S.L[4].

The cost of applying the test, estimated in €2.75, was calculated using the annual gross salary of the staff who performs the test and the number of hours they employ.

The costs of colonoscopy, biopsy, and polypectomy are €173, €75 and €341. They were taken from (López-Bastida et al., 2010), and updated according to the inflation data published by the Spanish National Institute of Statistics.

The costs of CRC treatment in Spain for stages 0, I, II, III and IV were €6,573, €20,298, €28,251, €36,894 and €27,001, respectively (Corral et al., 2015).

[4] http://www.boa.aragon.es/cgi-bin/EBOA/BRSCGI?CMD=VERLST&BASE=BOLE&DOCS=1-33&SEC=IMPRESION&SEPARADOR=&&PUBL=20131217

Policy	Cost	Effectiveness
Start screening at 50 years	€2,375	22.843 QALYs
Start screening at 60 years	€2,647	22.762 QALYs
No screening	€2,910	22.696 QALYs

Table 7.2: Cost-effectiveness results for the reference case.

7.3 Results

7.3.1 Reference-case analysis

As mentioned above, we have evaluated three different screening strategies with FOBT: starting the screening at 50 years, starting at 60, and not performing the screening. The model has a temporal horizon of 70 cycles, that is, until the cohort reached 100 years of age. In the reference case, a discount rate of 3% per year was applied for both costs and effectiveness (Weinstein et al., 1996).

Table 7.2 shows the cost and effectiveness of each policy. We can see that in the reference case, the cost of starting the screening at 50 is lower than the cost of starting at 60, which is in turn lower than that of no screening. Additionally, starting the screening at 50 yields more QALYs than starting at 60, which is in turn more effective than no screening. Therefore, in the reference case, starting the screening at 50 years dominates the other policies, i.e. it is cost-saving. This means that the cases of CRC avoided or detected in an early state lead to savings grater than the cost of starting the screening at a young age.

7.3.2 Sensitivity analysis

Also in this study we have modeled the uncertainty by assigning second-order probabilities to the parameters. The uncertainty of each probability has been modeled through a Beta distribution with a sample size $(\alpha + \beta)$ of 100 individuals. Uncertainty about costs has been modeled using Gamma distributions with a standard deviation of 10% over the reference value $\Gamma(\mu, \frac{1}{10}\mu)$. Even though Briggs et al. (2006) recommended the use of a standard deviation of 20%, which is the value we used in Chapter 6, in the

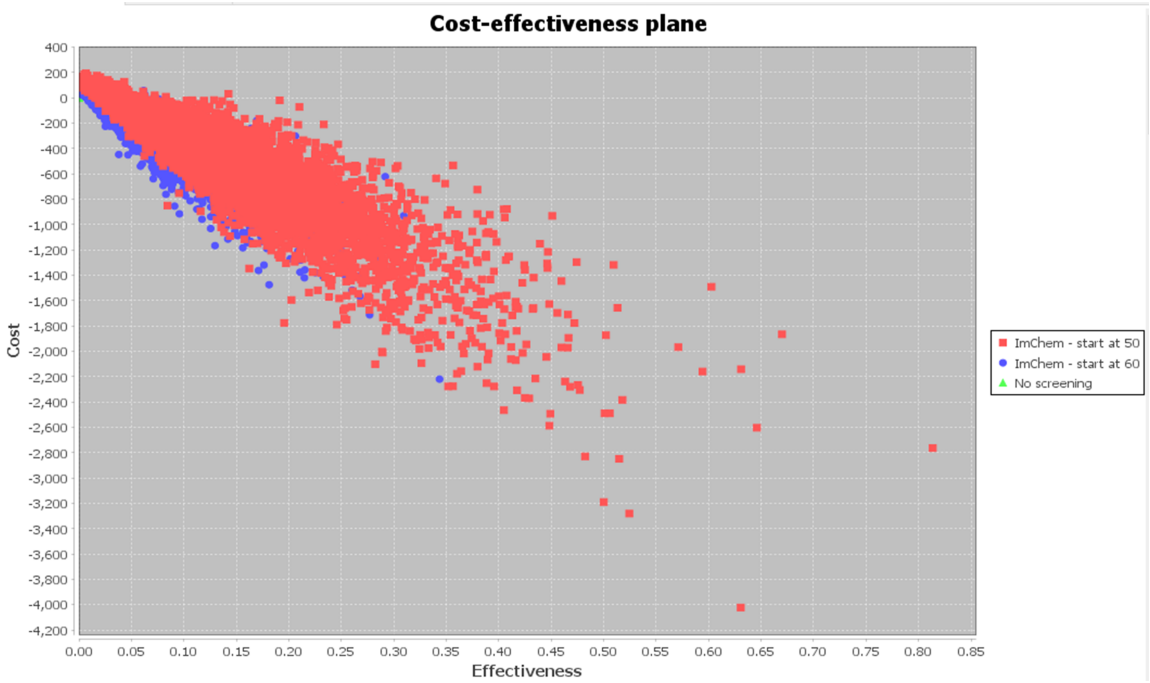


Figure 7.5: Cost-effectiveness scatter plot of the policies relative to no-screening. Each point represent a simulation measuring the costs in €and the effectiveness in QALYs. The origin of coordinates corresponds to no screening.

CRC model we assume a lower uncertainty on the costs because we obtained them from public official lists.

Using these second-order probability distributions we made 10,000 Monte Carlo simulations. Figure 7.5 shows the scatter plot for the incremental values of cost and effectiveness of the two screening strategies, taking as reference the strategy of no screening. Most simulations have lower cost and greater effectiveness, but for some simulations the cost of screening is higher, which means that the savings due to prevention and early detection are smaller than the costs of FOBT and colonoscopy.

The acceptability curve in Figure 7.6 compares the strategy of starting the screening at 50 with no screening. We can see that the strategy has a 95% probability of being cost-saving. For a WTP of €30,000/QALY, the probability that the screening be cost-effective is virtually 100%.

If we compare the two screening strategies, start the screening at 50 has a 99% probability of being cost-effective with respect to start the screening at 60.

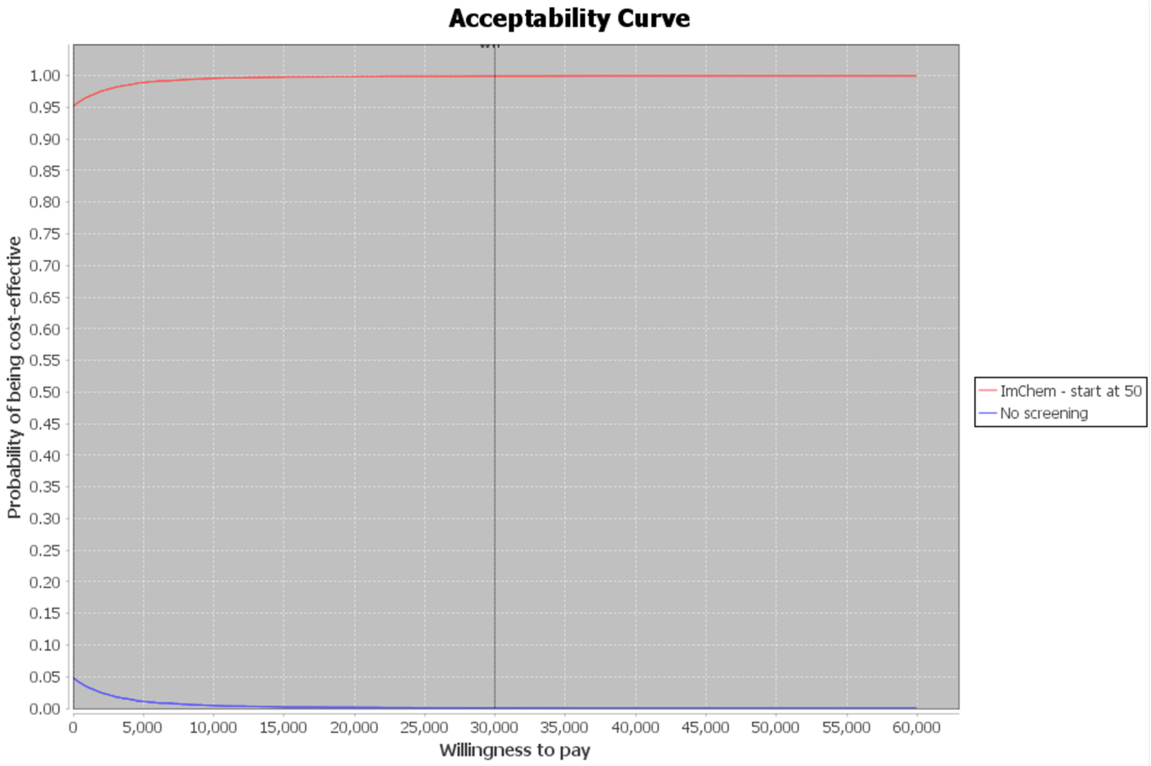


Figure 7.6: Acceptability curve comparing the strategies of starting at 50 versus no screening.

7.4 Discussion

Our study has compared three interventions: starting the screening at 50 years, starting at 60, and no screening. Although in Aragon it was decided to start the screening at 60 years, our study shows that starting it at 50, as recommended by the Spanish clinical practice guidelines, saves costs.

In our study we have only considered the immunochemical FOBT (iFOBT), while other authors (López-Bastida et al., 2010; Lejeune et al., 2014) also analyze the Guaiac test (gFOBT). Although this type of test was approved by the European Commission and recommended by the health authorities in some countries, such as France, it has been criticized because its low sensitivity and because it reacts with the non-human heme group in food (Macrae et al., 1982). In contrast, immunochemical tests, which are based on a specific analysis method for human hemoglobin, offer greater sensitivity and specificity without the need of a previous diet excluding meat or without peroxidases. Additionally, iFOBT only needs one sample to be reliable instead

the three needed by the gFOBT. For this reason, iFOBT presents better compliance. Furthermore, the samples can be analyzed automatically, which allows the process to be reproducible and subject to quality control. This explains why iFOBT, in spite of being more expensive, is cost-effective with respect gFOBT (López-Bastida et al., 2010; Lejeune et al., 2014). Hence we decided do not include the gFOBT in our model.

In our study we considered that the test is positive when the hemoglobin concentration is equal to or higher than 117 ng/mL. This is the cut-off point recommended by Hamza et al. (2013), used in the health system of Aragon, and established by default in the SENTIFIT-FOB Gold latex test. However, other authors argue that the most cost-effective cut-off point is 110 ng/mL (Chen et al., 2007). In addition, the concentration of hemoglobin in stool is related to the severity of the lesion, and therefore this result could be useful to stratify the risk and prioritize colonoscopies in patients with higher concentrations of blood in feces (Auge et al., 2014). Determining the most appropriate cut-off point for screening is a topic for future research.

Some authors consider the risk of death due to complications in diagnostic procedures, such as endoscopic perforation (López-Bastida et al., 2010). However, in our study we assumed that this probability is almost null, according to the results of other randomized trials (Faivre et al., 2004; Sobhani et al., 2011).

As we said, several parameters of our model were taken from the "Aragon study", mentioned in the first paragraph of this chapter, whose participants were people between 60 and 69 years old, since in Aragon it was decided to start the screening at 60 and advancing it progressively until reach the age of 50. The participation rate in the screening was 58.91%, very similar to that of previous studies (Tappenden et al., 2007; Sobhani et al., 2011). We assumed that the participation and compliance rates for patients aged from 50 to 59 are the same as those for patients from 60 to 69.

The other hypotheses in our model coincide with those commonly used in the literature. The evolution of the cohort begins at the age of 30, when the prevalence of adenomas and CRC is almost null. We assumed that progression to CRC occurs only from a preexisting adenoma (Levin et al., 2008; Sobhani et al., 2011). This is due to the lack of evidence related to the rate of de novo CRC. We also considered that 100% of the adenomas detected in colonoscopy are removed in the polypectomy.

The results of our model agree with those of previous cost-effectiveness studies of CRC screening, which concluded that FOBT is cost-effective (Tappenden et al., 2007; López-Bastida et al., 2010; Sobhani et al., 2011) and with a recent study (Arrospide et al., 2018) that indicates that it is not only cost-effective but also cost-saving in Spain.

Our model—as well as all the models used in the cost-effectiveness studies of our group—is available on the Internet so that other researchers can examine it, check the results and modify it—for example, changing the prevalence of CRC to adapt the model to other regions, varying the sensitivity and specificity values of the tests, introducing other screening patterns, etc. As we have not found any public available CRC screening model, it has not been possible to investigate all the reasons for the differences in costs and effectiveness values obtained.

7.5 Conclusions

According to our analysis, biennial CRC screening with the iFOBT dominates the alternative of no screening; i.e. it obtains greater effectiveness at a lower cost. This is mainly because the detection and removal of polyps increases life expectancy and avoid expensive treatments. If the screening start at the age of 50, the gain in effectiveness and the economic savings are greater than if it starts at 60. According to the PSA, the probability of these conclusions is close to 100% for the WTP of €30,000/QALY, which is accepted as a consensus of experts as the value for the Spanish health system.

The model used in this study is available on the internet, so that it can be examined and reused by other researchers.

Part V

Conclusion

8

Conclusions

"Everything should be made as simple as possible, but not simpler."

Albert Einstein

8.1 Main contributions

In this thesis we have shown the usefulness of Probabilistic Graphical Models (PGMs) for decision analysis. In particular we addressed three methodological problems related to the evaluation of cost-effectiveness models and applied them to two medical problems.

In Chapter 4 we have compared different approaches for minimizing the error introduced by the discretization of time in Markov models. There has been a lot of controversy about how to interpret and apply half-cycle correction, and whether it should be replaced by the life-table method. More recently were proposed within-cycle corrections, based on numerical-integration techniques, which were a good starting point to address the shortcomings of standard approaches. However, this methods are less accurate when the model has discontinuities. We proved that building a new model averaged at the points of discontinuity yields much more accurate results.

In Chapter 5 we have also studied how to perform Cost-Effectiveness Analysis (CEA) with PGMs. The Research Center for Intelligent Decision-Support Systems (CISIAD) is the only research group that has used probabilistic graphical models to model and analyze numerous medical problems, despite of that, their application for CEA is small yet. Although most studies are based on decision trees or Markov models, PGMs have several advantages: they have compact representations, can solve complex problems, and are easier to understand because are based on causal graphs that summarize the structure of the model. Markov Influence Diagrams (MIDs) allow temporal reasoning of models with large horizons. The existing cost-effectiveness algorithms for MIDs only evaluate models with two criteria and one decision. In Section 5.2 we have developed the algorithms that can evaluate MIDs with several criteria and any number of decisions, with findings between decisions.

Using Decision Analysis Networks (DANs) it is possible to solve many asymmetric decision problems, for example, those involving unordered or partially-ordered decisions, which cannot be modeled with traditional techniques. Like IDs, DANs can model some large problems for which a DT would contain millions of branches, but do not require a total ordering of the decisions. This way DANs combine the flexibility of DTs with the compactness of IDs. When we began this thesis, DANs could only evaluate

unicriterion problems. In Section 5.3 we have presented cost-effectiveness algorithms for DANs.

In Chapter 6 we have analyzed the cost-effectiveness of the pediatric Bilateral Cochlear Implantation (BCI) in Spain. We proved that this intervention is clearly cost-effective in Spain even if the system took over all the maintenance costs, which in Spain are partially covered by users and their families. Under the current policy, the probability of being cost-effective is almost 100%, while if the system covers maintenance costs, it would be 99.98%. This study has been one of the most satisfying parts of my thesis because of its social impact. The preliminary work conducted by our research group, which included a thorough review of the literature, contributed to convincing the Spanish Ministry of Health that it is clearly cost-effective. Thus Spain became the first country in the world—to the best of our knowledge—to include BCI for both children and adults in the portfolio of health services (cf. Orden SSI/1356/2015, de 2 de julio), but several regional governments still refused to cover it in practice, even for newborns. After obtaining the results of our model, we wrote a detailed report, and submitted it to the Ministry of Health and to 11 regional health departments. In May 2018 the Ministry of Health confirmed that BCI must be covered in Spain and the health departments that had been reluctant to finance this intervention, finally approved it: Andalusia in September 2018 and Catalonia in March 2019. We are satisfied because our research on medical AI has made a difference on the life of patients.

ColoRectal Cancer (CRC) is one of the most frequent causes of morbidity and mortality in the world. The goal of screening programs, is to detect the disease at an early stage or even prevent it. In Chapter 7, we have compared three interventions: starting the screening with fecal occult blood test at 50 years, starting at 60, and no screening. Although in Aragon it was decided to start the screening at 60 years, our study shows that starting at 50, as recommended by the Spanish clinical practice guidelines, saves costs. It has a 95% of probability of being cost-saving and almost a 100% of being cost-effective.

These models and algorithms, like all those built by our group, are publicly available, so that any researcher can reproduce our results, perform new analyses or adapt our models to their needs. In spite the reproducibility is one of the basic tenets

of the scientific method (Peng, 2011; Repko and Szostak, 2016), the models used for CEA are almost never publicly available.

During this thesis it has been necessary to implement several the algorithms and functionalities. We added multiple new functionalities to OpenMarkov, such as the definition and evaluation of multi-criteria networks, the integration of Markov Processes with Atemporal Decisions (MPADs) and influence diagrams, the representation of the temporal evolution of a variable or the graphical representation of the results obtained in cost-effectiveness and sensitivity analyses (see App. C). I have also devoted a large amount of time to maintenance tasks, such designing new tests, debugging, and documenting OpenMarkov, because, rather than just building a research tool for our group, we have built an open-source program that can be used by any researcher in any country.

8.2 Publications

8.2.1 Journal publications

1. F. J. Díez, M. Yebra, I. Bermejo, M. A. Palacios-Alonso, M. Arias, M. Luque, and J. Pérez-Martín (2017). "Markov influence diagrams: A graphical tool for cost-effectiveness analysis." In: *Medical Decision Making* 37, pp. 183–195
2. J. Pérez-Martín, M. A. Artaso, and F. J. Díez (2017a). "Cost-effectiveness of pediatric bilateral cochlear implantation in Spain." In: *The Laryngoscope*
3. J. Pérez-Martín, I. Bermejo, and F. J. Díez (2019). "Evaluation of Markov models with discontinuities." In: *Medical Decision Making*

8.2.2 Conference publications

1. M. Lalana, J. Pérez-Martín, A. Fontán, and F. J. Díez (2016b). "Cost-effectiveness of colorectal cancer screening in Aragon (Spain)." In: *16th Biennial European Conference of the Society for Medical Decision Making*. London, UK

2. J. Pérez-Martín, M. A. Artaso, and F. J. Díez (2016). "Cost-utility analysis of bilateral cochlear implants in children." In: *Binaural Hearing with Hearing Implants: A Tribute to 20 Years of Bilateral Cochlear Implantation*. München, Germany

3. M. Lalana, A. Fontán, F. J. Díez, J. Pérez-Martín, A. Tapia, J. Millastre, M. M. Larrea, and M. Sánchez (2016a). "Estudio coste-efectividad del cribado de cáncer colorrectal en nuestro sector durante los años 2014-2015." In: *X Congreso Nacional del Laboratorio Clínico*. Zaragoza, España

4. M. Arias, M. A. Artaso, I. Bermejo, F. J. Díez, M. Luque, and J. Pérez-Martín (2017a). "Advanced algorithms for medical decision analysis. Implementation in OpenMarkov." In: *Proceedings of the 16th Conference on Artificial Intelligence in Medicine (AIME 2017)*. Vienna, Austria

5. M. Arias, M. Luque, J. Pérez-Martín, and F. J. Díez (2017b). "Cost-effectiveness analysis with decision analysis networks." In: *Annual Meeting of the Society for Medical Decision Making*. Pittsburgh, PA

6. J. Pérez-Martín, I. Bermejo, and F. J. Díez (2017b). "The problem of discontinuities in the evaluation of Markov models." In: *Annual Meeting of the Society for Medical Decision Making*. Pittsburgh, PA

7. F. J. Díez, I. París, J. Pérez-Martín, and M. Arias (2018c). "Teaching Bayesian networks with OpenMarkov." In: *Proceedings of the Ninth European Workshop on Probabilistic Graphical Models (PGM'18)*. Prague, Czech Republic

8. F. J. Díez, M. Luque, J. Pérez-Martín, and M. Arias (2018b). "Research on medical decision analysis at the CISIAD, UNED." in: *Proceedings of the XVIII Conference of the Spanish Association for Artificial Intelligence (CAEPIA-2018)*. Granada, Spain

9. M. Arias, J. Pérez-Martín, M. Luque, and F. J. Díez (2019). "OpenMarkov, an open-source tool for probabilistic graphical models." In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*

8.3 Ongoing and future work

Our group is currently working on the project "Cost-effectiveness analysis with decision analysis networks" supported by a national grant (ref. TIN2016-77206- R). We are collaborating with a multidisciplinary team that involves researchers from Universidad Complutense de Madrid, Hospital de la Paz (Madrid) and the Spanish group HM Hospitals. We are working on an extended model for the CRC (see ch. 7) in which we are using a DAN to analyze more complex policies. In the current form the new model will be able to analyze seven decisions, each one with two or more alternatives. We are thus analyzing 1,296 scenarios, performing CEA for all of them and comparing the ICER of all these alternatives. This model might improve the current guidelines on CRC in Spain.

In the future, taking as a basis our study of the CEA of pediatric BCI in Spain, we want to adapt our model to other countries. We believe that our model can be easily adapted to other countries making only slight modifications, helping other people to demonstrate if the BCI is cost-effective in their countries.

We want to measure the impact of discontinuities in real-world problems with uncertainty. The recommendations of Briggs et al. (2006) on how to address uncertainty, followed by most of the experts in this field, could have such a high impact that the error introduced at selecting a half-cycle correction method would be negligible. We want to study this effect in models where uncertainty is based those recommendations and also in models in which the uncertainty is modeled through real data.

Another line for future research is the development of more efficient algorithms for DANs, including parallel implementations and the design of sensitivity analysis options and explanation facilities similar to those existing for ID (Lacave et al., 2007). Also we want to develop Markov DANs, which would combine the advantages of DANs with those of Markov IDs (Díez et al., 2017) and try to compact the interventions returned by the algorithms (Prada, 2014).

Finally, we will continue working on OpenMarkov, adding new functionalities and debugging the current issues, looking for release the first official stable version.

Part VI

Appendices

A

Resumen (Summary in Spanish)

Dado que los recursos de los sistemas de salud son limitados, la evaluación económica de las técnicas y tratamientos médicos es cada vez más importante. Los árboles de decisión y los modelos de Markov son las herramientas más utilizadas para el análisis de coste-efectividad, pero solo pueden resolver problemas relativamente pequeños. Los modelos gráficos probabilistas, como las redes bayesianas y los diagramas de influencia, se han utilizado en inteligencia artificial para la representación y explicación del conocimiento, especialmente en medicina, pero únicamente en problemas unicriterio. En los últimos años, el Centro de Investigación sobre Sistemas Inteligentes de Ayuda a la Decisión (CISIAD) en la UNED, ha desarrollado nuevos algoritmos para realizar análisis de coste-efectividad con árboles de decisión y diagramas de influencia. También ha propuesto dos nuevos tipos de modelos gráficos probabilistas: los diagramas de influencia markovianos, que amplían los diagramas de influencia para incluir razonamiento temporal y las redes de análisis de decisiones, que pueden modelar y evaluar problemas con asimetrías como restricciones o decisiones parcialmente ordenadas.

Esta tesis aborda tres problemas metodológicos relacionados con la evaluación de modelos de coste-efectividad.

Primero, hay diferentes métodos para reducir el error introducido por la discretización del tiempo en los modelos de Markov. En general, las técnicas de integración numérica dan resultados más precisos que los enfoques tradicionales, como la corrección de medio ciclo, pero pueden conducir a un error mayor cuando el modelo tiene discontinuidades, por ejemplo, cuando se un tratamiento caro se retira después de aplicarlo durante algún tiempo. Hemos probado que, la construcción de un nuevo modelo promediado en los puntos de discontinuidad produce resultados mucho más precisos.

Segundo, los algoritmos de coste-efectividad existentes para los diagramas de

influencia markovianos solo eran capaces de evaluar modelos con dos criterios y una decisión. En esta tesis, he desarrollado un nuevo algoritmo de coste-efectividad que permite evaluar modelos con varios criterios y cualquier número de decisiones con hallazgos entre las decisiones.

Tercero, las redes de análisis de decisiones solo podían evaluar problemas unicriterio. En colaboración con otros miembros de la CISIAD, he ampliado los algoritmos desarrollados para los diagramas de influencia markovianos, para poder realizar análisis de coste-efectividad en redes de análisis de decisión.

He aplicado los diagramas de influencia markovianos a la evaluación económica de dos intervenciones médicas. Nuestro análisis del implante coclear pediátrico en España ha demostrado que es coste-efectivo con respecto al implante coclear unilateral. El modelo para el cribado del cáncer colorrectal con test inmunoquímico de sangre oculta en heces mostró que permite ahorrar costes con respecto a la estrategia de no cribado.

B

Evaluation of the DAN for the 2-test problem

In Section 5.3.2 we have used the 2-test DAN to explain the evaluateDAN algorithm, which recursively decomposes a DAN until it contains no decisions. In this section we offer the same explanation in much more detail.

B.1 Decomposition of the DAN

Every invocation of the Algorithm 5.2 takes two arguments: a DAN and a set of findings, denoted by **e** (evidence). The first invocation always takes the original network and no evidence. In our example, the DAN has no always-observed variable, so the assignment in line 4 makes $\mathbf{O} \leftarrow \varnothing$. The network has three decisions: D_A, D_B, and T. In principle any of them might be made first, but the algorithm detects that T cannot reveal any information, so it must not be the first one; in line 23 it makes $\mathbf{D}_I \leftarrow \{D_A, D_B\}$ and then decomposes the original network into two DANs, as shown in Figure 5.9: in one of them D_A is prioritized by drawing the links $D_A \rightarrow D_B$ and $D_A \rightarrow T$; in the other, D_B is prioritized. Line 28 adds a meta-decision variable OD (for "order of the decisions"), indicating which decision will be made first. If we modify Algorithm 5.2 in order to explicitly build the DT, this variable would be the root (see Fig. 5.10); in other examples there might be several meta-decisions.

The evaluation of the network containing the link $D_A \rightarrow D_B$ arrives at line 16 and generates two new DANs, one for $+d_A$ (do test A) and one for $\neg d_A$ (do not perform test A), which no longer contain the decision D_A. This decomposition corresponds to the branches $+d_A$ and $\neg d_A$ in Figure 5.10. In the new DANs the node D_A disappears

because this decision has been made.

In the DAN for $+d_A$, the method instantiate marks R_A as always-observed. For this reason it has a red oval around it, which means that its value is known when making the next decisions. The tables $P(r_A|d_A)$ and $c_A(d_A)$ have been projected for $+d_A$. Line 4 of Algorithm 5.2 makes $\mathbf{O} \leftarrow \{R_A\}$ and line 12 invokes evaluateDAN twice: first with $\mathbf{e} = \{+r_A\}$ and then with $\mathbf{e} = \{\neg r_A\}$. In both cases the DAN is decomposed for D_B—as it was previously decomposed for D_A—and then for R_B, T, and X. The upper node CE in Figure 5.10 corresponds to an invocation of evaluateDAN with a DAN in which all the decisions have been removed and every chance node has been assigned a value, $\mathbf{e} = \{+r_A, +r_B, +x\}$.

In the DAN for $\neg d_A$ (see again Fig. 5.9), the node R_A has disappeared because the two values of this variable, "positive" and "negative", are incompatible with $\neg d_A$, due to the total restriction for link $D_A \rightarrow R_A$. This DAN is subsequently decomposed as in the previous case, and also the DAN containing the link $D_B \rightarrow D_A$ is evaluated in the same way.

B.2 Probabilities and CEPs returned

Every recursive call to evaluateDAN (Algorithm 5.2) returns a probability, i.e., a single number between 0 and 1, and a CEP. The call corresponding to the upper CE node in Figure 5.10 returns the probability $P(\mathbf{e}) = P(+x, +r_A, +r_B) = P(+x) \cdot P(+r_A \mid +x) \cdot P(+r_B \mid +x) = 0.14 \cdot 0.78 \cdot 0.90 = 0.09828$, computed from the conditional probabilities of the chance nodes, just as in the case of BNs. The CEP for this node consists of a single interval, $(0, +\infty)$, which implies that every leaf node has only one value of cost and one of effectiveness, just as in the standard CEA algorithm for DTs. The cost and the effectiveness for this CEP is the sum of the values of the corresponding nodes: $c = c_A(+d_A) + c_B(+d_B) + c_T(t_2) = 18 + 150 + 70,000 = 70,168$ and $e = e(+x, t_2) = 6.5$—let us remember that decisions that led to this DAN are $D_A = +d_A$, $D_B = +d_B$, and $T = t_2$. The intervention is empty because no decision node has yet been evaluated.

For the other U node shown in Figure 5.10 we have $P(\mathbf{e}) = P(\neg x, +r_A, +r_B) =$

$P(\neg x) \cdot P(+r_A \mid \neg x) \cdot P(+r_B \mid \neg x) = 0.86 \cdot 0.09 \cdot 0.07 = 0.005418$, $c = c_A(+d_A) + c_B(+d_B) + c_T(t_2) = 70,168$, $e = e(\neg x, t_2) = 9.3$, and the empty intervention.

The algorithm evaluateDAN steps back to node X in that figure. It corresponds to a call in which $\mathbf{e} = \{+r_A, +r_B\}$. Line 13 of Algorithm 5.2 computes $P(+r_A, +r_B) = P(+x, +r_A, +r_B) + P(\neg x, +r_A, +r_B) = 0.103698$ and line 14 computes $P(+x \mid +r_A, +r_B) = 0.94775$ and $P(\neg x \mid +r_A, +r_B) = 0.05225$. These are the probabilities for the two branches outgoing from node X in Figure 5.10. When averageCEP (Algorithm B.1) is invoked in the next line, these probabilities are used to average the cost and the effectiveness in the resulting CEP, which still has an empty intervention and no threshold because no decision node has yet been evaluated.

The recursion steps back to node T in that figure. It corresponds to a call to evaluateDAN with a network that still contained the decision node T. The "for" loop in line 17 returns three probability values and three CEPs, one from each branch of T in the tree. The three probabilities are identical, because $P(\mathbf{e}) = P(+r_A, +r_B)$ and the results of the tests do not depend on the therapy selected after doing them. So line 20 can arbitrarily select any of the $P_d(\mathbf{e})$'s. The three CEPs are defined on the same interval, $(0, +\infty)$, but each one has a different cost, effectiveness, and intervention. Then line 21 invokes the method optimalCEP (Algorithm B.2), which performs a deterministic analysis (with Algorithm B.3) for the only interval of the input CEPs and returns a CEP that, for the numerical parameters of this model, has three intervals.

Node R_B in Figure 5.10 corresponds to the top right DAN in Figure 5.8. The "for" loop in line 10 receives two probabilities, $P(+r_A, +r_B)$ and $P(+r_A, \neg r_B)$. Line 14 computes $P(+r_B \mid +r_A) = 0.5557$ and $P(\neg r_B \mid +r_A) = 0.4443$, which are the probabilities of its outgoing branches. The CEP for $+r_A$ contains two thresholds, €7,551.50/QALY and €21,385.50/QALY, while the CEP for $\neg r_A$ contains one, €70,923.70/QALY. The method averageCEP, invoked in the next line, computes the average cost and the average effectiveness for each of the four intervals. In the first one, i.e., when $\lambda <$ €7,551.50/QALY, the optimal intervention is not to do test B and not to apply any therapy, even though A has given a positive result. For the other intervals, test B is done. The optimal intervention for the second interval is: "if B is negative, apply no therapy; if it is positive, therapy 1 is applied". The optimal intervention for the

third interval is similar, with therapy 2 instead of therapy 1. In the fourth interval, a positive result of B leads to therapy 1 and a negative result to therapy 2.

The recursion continues all the way back to the root of the trees, which—as we said above—corresponds to a call to evaluateDAN with the original DAN no evidence. The "for" loop in line 24 returns two probabilities, which are both equal to 1, and two CEPs, one for the DAN in which the decision D_A is made before D_B and another one for the DAN in which D_B is made first. Line 28 creates the auxiliary variable OD, i.e., the meta-decision that determines, for each λ-interval, which decision must be made first.

B.3 Auxiliary algorithms

In this section we present some auxiliary algorithms invoked by the main method, Algorithm 5.2. They were introduced to evaluate DTs with embedded decision nodes (Arias and Díez, 2011) and later used for evaluating IDs (Arias and Díez, 2015). We reproduce them here for the sake of completeness.

The method instantiate, invoked in lines 11 and 18, takes as arguments a DAN, a variable X, and a value x, and generates a new network, DAN_x, as follows:

1. create DAN_x as a copy of the DAN;
2. if x reveals a variable Y that is not a descendant of any decision (other than X), then declare Y as observed in DAN_x; if Y is a descendant of some decisions (other than X), $\{D'_1, \ldots, D'_n\}$, draw a link $D'_i \to Y$ for $i \in \{1, \ldots, n\}$—if it did not exist—and declare that D'_i reveals Y unconditionally;
3. if there is a total restriction (x, Y), remove Y; then remove recursively all the chance and utility nodes that are children of Y;
4. if there is a restriction (x, y), remove y from the domain of Y;
5. if a chance node Y is a child of X, project the table $P(y|pa(Y))$ or $u_Y(pa(Y))$ by making $X = x$;
6. if X is a decision, remove node X and its outgoing links.

Line 15 of Algorithm 5.2 invokes the method averageCEP, given by Algorithm B.1. Its input consists of a variable X having m values (states), a probability distribution $P(x)$, and m CEPs. It first gathers all the thresholds of the input CEPs and computes, for each interval, the average cost, the average effectiveness, and the intervention, taken the cost, effectiveness, and interventions of the input CEPs for those intervals. Again, an example can be found in (Arias and Díez, 2011).

Line 21 of Algorithm 5.2 invokes the method optimalCEP, given by Algorithm B.2. Its input consists of a decision D having m values (options) and m CEPs. It first gathers all the thresholds of the input CEPs and performs, within each interval, a deterministic CEA, which may in turn generate new thresholds. Those thresholds that lie within the interval analyzed are added to the set of CEPs, Θ; the others are

Algorithm B.1: Weighted average of CEPs

function : `averageCEP`

input : X – a chance variable whose domain is $\{x_1, \ldots, x_m\}$,
$P(x_j)$ – a probability distribution for X, and
$\{Q_1, \ldots, Q_m\}$ – a set of CEPs

output : *CEP* – a cost-effectiveness partition

1 $\Theta \leftarrow \bigcup_{j=1}^{m} \Theta_j$
2 $n \leftarrow \text{ card}(\Theta) + 1$
3 **for** $i \leftarrow 1$ **to** n **do**
4 $\quad c_i \leftarrow \sum_{j=1}^{m} P(x_j) \cdot cost_{Q_j}(\theta_i)$
5 $\quad e_i \leftarrow \sum_{j=1}^{m} P(x_j) \cdot \mathit{eff}_{Q_j}(\theta_i)$
6 $\quad I_i \leftarrow$ "If $X = x_1$, then $interv_{Q_1(\theta_i)}$; if $X = x_2$, then $interv_{Q_2(\theta_i)}$..."
7 **return** $((\theta_1, \ldots, \theta_{n-1}), (c_0, \ldots, c_n), (e_0, \ldots, e_n), (I_0, \ldots, I_n))$

discarded. For each of the new intervals, the algorithm selects the value d_k of D that maximizes the NMB. It determines the cost and the effectiveness for that interval, just as when selecting the optimal branch outgoing from a decision node in a tree. The optimal intervention for that interval begins by selecting $D = d_k$ and continues with the partial intervention that the k-th CEP determined for that interval. Finally, the algorithm eliminates the thresholds that are not necessary, i.e., those that separate intervals having the same cost, effectiveness, and intervention. Those thresholds were generated in branches of the tree that later turned out to be suboptimal. An example can be found in (Arias and Díez, 2011).

The deterministic CEA performed in line 4 of this algorithm might be done with the traditional method, which consists in keeping the non-dominated interventions and discarding those dominated by others (single dominance or extended dominance) (Weinstein et al., 1980). A slightly more efficient method, introduced in (Arias and Díez, 2011), is given by Algorithm B.3. It begins by selecting the intervention with the lowest cost and then the one with the lowest ICER among those with higher effectiveness, and so on.

Finally, the method prioritize, invoked in line 25 of Algorithm 5.2 with a DAN and a decision node D as arguments, is implemented by just drawing links to the other

Algorithm B.2: Optimal CEP

function : `optimalCEP`

input : D – a decision node whose domain is $\{d_1, \ldots, d_m\}$ and $\{Q_1, \ldots, Q_m\}$, a set of m CEPs with $Q_j = (\Theta_j, C_j, E_j, I_j)$

output : CEP – a cost-effectiveness partition

1 $\Theta \leftarrow \bigcup_{j=1}^{m} \theta_j$
2 $n \leftarrow$ card(Θ)+1
3 **for** $i \leftarrow 1$ **to** n **do**
4 perform a deterministic CEA analysis within the i-th interval
5 add to Θ the new thresholds that belong to this interval
6 $n \leftarrow$ card(Θ)+1
7 **for** $i \leftarrow 1$ **to** n **do**
8 $k \leftarrow \arg\max_j \ NMB(\theta_i)$
9 $c_i \leftarrow cost_{Q_k}(\theta_i)$
10 $e_i \leftarrow eff_{Q_k}(\theta_i)$
11 $I_i \leftarrow$ "$D = d_k$; $interv_{Q_k(\theta_i)}$..."
12 fuse contiguous intervals that have the same intervention, the same cost, and the same effectiveness.

candidates to be the first decision, i.e., the nodes in $\mathbf{D}_I$.

Algorithm B.3: Deterministic cost-effectiveness analysis

function : `deterministicCEA`

input : a set of interventions $\{I_1, \ldots, I_m\}$

output : *CEP* – a cost-effectiveness partition

1 $\sigma(0) \leftarrow \arg\min_i \ cost(I_i)$

2 $R_0 \leftarrow \{i \mid I_i \in I \wedge \mathit{eff}(I_i) > \mathit{eff}(I_{\sigma(0)})\}$

3 $i \leftarrow 1$;

4 **while** $R_{i-1} \neq \varnothing$ **do**

5 $\quad \sigma(i) := \arg\min_{j \in R_i} \ ICER(I_{\sigma(i-1)}, I_j)$

6 $\quad \theta_i \leftarrow \min_{j \in R_i} \ ICER(I_{\sigma(i-1)}, I_j)$

7 $\quad c_i \leftarrow cost(I_{\sigma(i)})$

8 $\quad e_i \leftarrow \mathit{eff}(I_{\sigma(i)})$

9 $\quad R_i \leftarrow \{j \mid j \in R_{i-1} \wedge \mathit{eff}(I_j) > \mathit{eff}(I_{\sigma(i)})\}$

10 $\quad i \leftarrow i + 1$

11 **return** $((\theta_1, \ldots, \theta_{n-1}), (c_0, \ldots, c_n), (e_0, \ldots, e_n), (I_0, \ldots, I_n))$

C

OpenMarkov

OpenMarkov is an open-source tool for building and evaluating several types of probabilistic graphical models. It implements several learning algorithms and can perform cost-effectiveness analysis and different types of sensitivity analysis. It is developed at the Research Center for Intelligent Decision-Support Systems (CISIAD) and constitutes a toolbox for researchers who wish to use probability-based models in articial intelligence. OpenMarkov has been used for research and teaching in more than 30 countries, including the Massachusetts Institute of Technology, Los Alamos National Laboratory, the National Oceanic and Atmospheric Administration, and several multinational companies.

The implementation of the cost-effectiveness algorithms developed in this thesis have required several changes in OpenMarkov's organization.

First of all, we made a deep re-factorization of the code in order to simplify the use of its Application Programming Interface (API), in accordance with the architectural pattern Model-View-Controller, separating the data, the logic, and the graphical representation. After analyzing OpenMarkov's tasks and how they worked, we were able to clean some unnecessary and/or duplicated code structures, by establishing the tasks and factoring out their common elements.

The second main contribution to OpenMarkov's coded was analyzing and re-organizing the 50 Maven modules that composed it. Some of them were using old versions of other modules, due to the absence of a well-defined structure. We cleaned this structure, removing unnecessary dependencies and revising the responsibilities of each module. The new structure allowed us to recover the continuous-integration practices and helped us to improve the documentation of the software.

We added multiple new functionalities, such as the encoding and evaluation of multi-criteria networks, the representation of the temporal evolution of a variable,

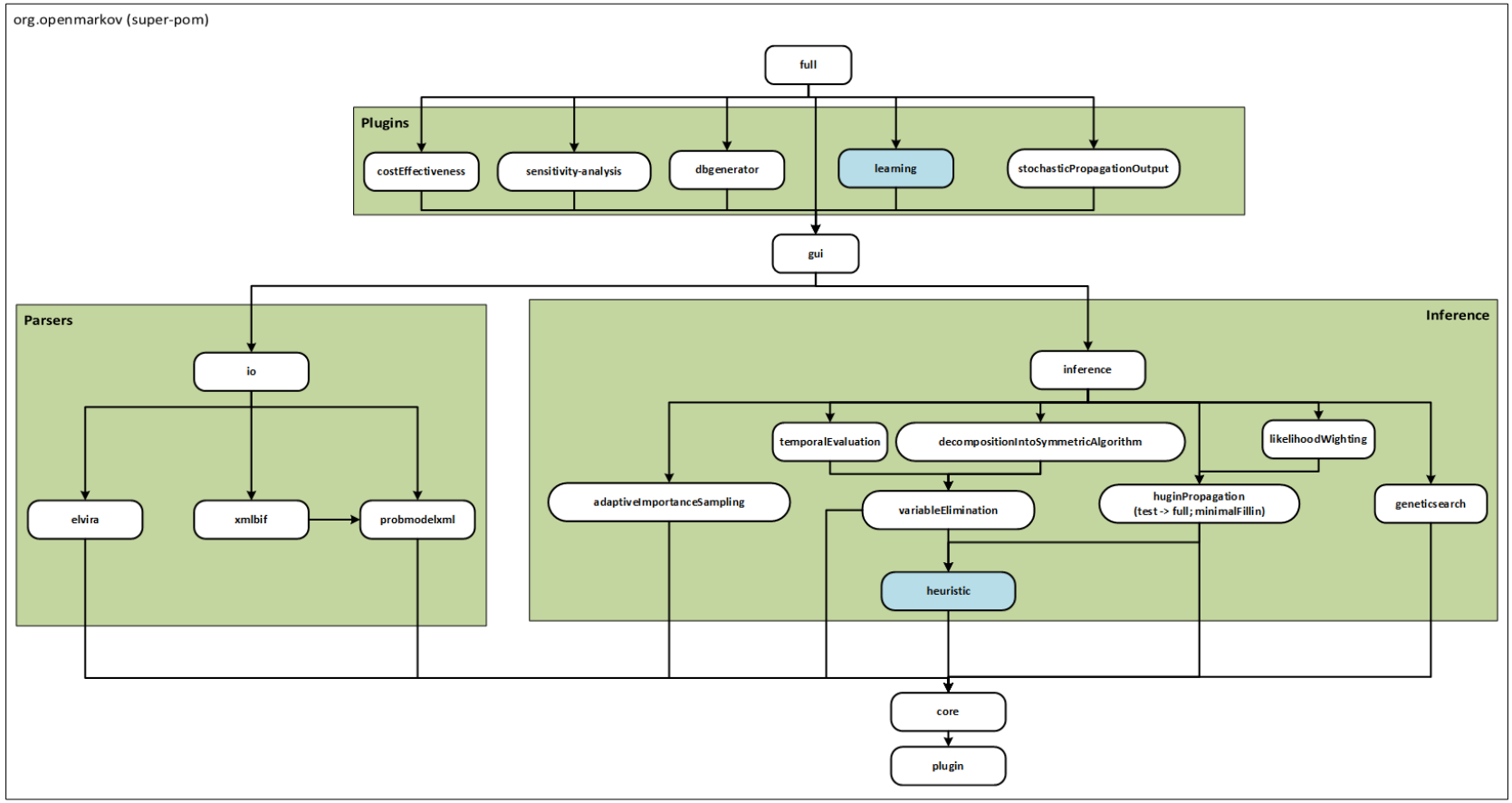


Figure C.1: Maven organization of OpenMarkov. Each blue node represents a group of several Maven modules.

or the graphical representation of the CEA and the PSA. It has also been necessary to perform some routine maintenance tasks, such as designing new tests, debugging issues, improving the documentation of OpenMarkov.

The networks described in chapters 6 and 7, were built in OpenMarkov and are publicly available in the website `www.probmodelxml.org`. Because ProbModelXML is the native encoding format of OpenMarkov's networks, both medical applications can be opened and evaluated with this open-source software.

During this thesis, we presented OpenMarkov in several peer-reviewed workshops and conferences (Arias et al., 2017a; Díez et al., 2018c; Arias et al., 2019). For a detailed view of OpenMarkov capabilities please see the wiki[1], the tutorial[2], or the video demonstration[3].

[1] `http://wiki.openmarkov.org`
[2] `http://openmarkov.org/docs/tutorial/`
[3] `http://openmarkov.org/ijcai-19/`

Bibliography

Allison, J. E., I. S. Tekawa, L. J. Ransom, and A. L. Adrain (1996). "A comparison of fecal occult-blood tests for colorectal-cancer screening." In: *New England Journal of Medicine* 334, pp. 155–160 (cit. on p. 120).

Amorín Calzada, M. J. (2008). *Mapa sanitario de la comunidad de Aragón 2007*. Tech. rep. Dirección General de Planificación y Aseguramiento. Departamento de Salud y Consumo. URL: `https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/docs/ResumenEjecutivoCribadoCancer.pdf` (cit. on p. 120).

Andreassen, S., M. Woldby, B. Falck, and S. K. Andersen (1987). "MUNIN — A causal probabilistic network for interpretation of electromyographic findings." In: *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI'87)*. Ed. by J. P. McDermott. Milan, Italy: Morgan Kaufmann, pp. 366–372 (cit. on p. 13).

Arias, M., M. A. Artaso, I. Bermejo, F. J. Díez, M. Luque, and J. Pérez-Martín (2017a). "Advanced algorithms for medical decision analysis. Implementation in OpenMarkov." In: *Proceedings of the 16th Conference on Artificial Intelligence in Medicine (AIME 2017)*. Vienna, Austria (cit. on pp. 50, 64, 133, 149).

Arias, M. and F. J. Díez (2011). *Cost-effectiveness analysis with sequential decisions*. Technical Report CISIAD-11-01. Madrid, Spain: UNED (cit. on pp. 37, 64, 70, 82, 143, 144).

Arias, M. and F. J. Díez (2014). "The problem of embedded decision nodes in cost-effectiveness decision trees." In: *Pharmacoeconomics* 32, pp. 1141–1145 (cit. on pp. 20, 21, 64, 74, 80).

Arias, M. and F. J. Díez (2015). "Cost-effectiveness analysis with influence diagrams." In: *Methods of Information in Medicine* 54, pp. 353–358 (cit. on pp. 37, 64, 70, 82, 143).

Arias, M., M. Luque, J. Pérez-Martín, and F. J. Díez (2017b). "Cost-effectiveness analysis with decision analysis networks." In: *Annual Meeting of the Society for Medical Decision Making*. Pittsburgh, PA (cit. on pp. 64, 133).

Arias, M., J. Pérez-Martín, M. Luque, and F. J. Díez (2019). "OpenMarkov, an open-source tool for probabilistic graphical models." In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (cit. on pp. 64, 133, 149).

Arrospide, A., I. Idigoras, J. Mar, H. de Koning, M. van der Meulen, M. Soto-Gordoa, J. M. Martinez-Llorente, I. Portillo, E. Arana-Arri, O. Ibarrondo, et al. (2018). "Cost-effectiveness and budget impact analyses of a colorectal cancer screening programme in a high adenoma prevalence scenario using MISCAN-Colon microsimulation model." In: *BMC Cancer*. 18, p. 464 (cit. on p. 126).

Artaso, M. and F. J. Díez (2016). *An empirical comparison of visual analog scales and three versions of the time trade-off. A study on cochlear implantation*. Technical Report CISIAD-16-01. Madrid, Spain: UNED (cit. on pp. 99, 102, 103, 106, 109).

Auge, J. M., M. Pellise, J. M. Escudero, C. Hernandez, M. Andreu, J. Grau, A. Buron, M. López-Cerón, X. Bessa, A. Serradesanferm, et al. (2014). "Risk stratification for advanced colorectal neoplasia according to fecal hemoglobin concentration in a colorectal cancer screening program." In: *Gastroenterology* 147, pp. 628–636 (cit. on p. 125).

Balkany, T., A. Hodges, F. Telischi, R. Hoffman, J. Madell, S. Parisier, B. Gantz, R. Tyler, R. Peters, and R. Litovsky (2008). "William House Cochlear Implant Study Group: position statement on bilateral cochlear implantation." In: *Otology & Neurotology* 29, pp. 107–108 (cit. on p. 91).

Barendregt, J. J. (2009). "The half-cycle correction: banish rather than explain it." In: *Medical Decision Making* 33, pp. 961–970 (cit. on pp. 48, 50, 55, 56, 60).

Barendregt, J. J. (2014). "The life table method of half-cycle correction: getting it right." In: *Medical Decision Making* 29, pp. 500–502 (cit. on pp. 48–50, 55, 60).

Bartlett, M. S. (1935). "Some aspects of the time-correlation problem in regard to tests of significance." In: *Journal of the Royal Statistical Society* 98, pp. 536–543 (cit. on p. 12).

Barton, G. R., P. C. Stacey, H. M. Fortnum, and A. Q. Summerfield (2006). "Hearing-impaired children in the United Kingdom, IV: Cost-effectiveness of pediatric cochlear implantation." English. In: *Ear and Hearing* 27, pp. 575–588 (cit. on pp. 92, 94).

Beck, J. R. and S. G. Pauker (1983). "The Markov process in medical prognosis." In: *Medical Decision Making* 3, pp. 419–458 (cit. on pp. 25, 48, 54).

Bess, F. H. (1986). "The unilaterally hearing-impaired child: a final comment." English. In: *Ear and Hearing* 7, p. 52 (cit. on p. 91).

Bichey, B. G. and R. T. Miyamoto (2008). "Outcomes in bilateral cochlear implantation." English. In: *Otolaryngology-Head and Neck Surgery* 138, pp. 655–661 (cit. on pp. 93–95, 109).

Bielza, C., M. Gómez, and P. P. Shenoy (2010). "Modeling challenges with influence diagrams: Constructing probability and utility models." In: *Decision Support Systems* 49, pp. 354–364 (cit. on p. 83).

Bielza, C., M. Gómez, and P. P. Shenoy (2011). "A review of representation issues and modelling challenges with influence diagrams." In: *Omega* 39, pp. 227–241 (cit. on pp. 20, 28).

Blalock, H. M. (1971). *Causal models in the social sciences*. Routledge (cit. on p. 13).

Bond, M., S. Mealing, R. Anderson, et al. (2007). *The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: A systematic review and economic model*. Report. Exeter, UK: Peninsula Technology Assessment Group (PenTAG) (cit. on pp. 92–94, 100, 103, 110).

Bond, M., S. Mealing, R. Anderson, J. Elston, G. Weiner, R. S. Taylor, M. Hoyle, Z. Liu, A. Price, and K. Stein (2009a). "The effectiveness and cost-effectiveness

of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model." In: *Health Technology Assessment* 13, pp. 1–330 (cit. on pp. 93, 94, 99).

Bond, M., K. Stein, S. Mealing, R. Anderson, J. Elston, G. Weiner, R. S. Taylor, M. Hoyle, Z. Liu, and A. Price (2009b). "The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: A systematic review and economic model." English. In: *Health Technology Assessment* 13, pp. 1–196 (cit. on pp. 98, 99, 108, 110, 111).

Braz, L. G., D. G. Braz, D. S. da Cruz, L. A. Fernandes, N. S. P. Módolo, and J. R. C. Braz (2009). "Mortality in anesthesia: A systematic review." In: *Clinics* 64, pp. 999–1006 (cit. on pp. 99, 100).

Briggs, A., K. Claxton, and M. Sculpher (2006). *Decision Modelling for Health Economic Evaluation*. New York: Oxford University Press (cit. on pp. 39, 49, 64, 102, 104, 122, 134).

Catalá-López, F., M. Ridao, A. Alonso-Arroyo, A. García-Altés, C. Cameron, D. González-Bermejo, R. Aleixandre-Benavent, E. Bernal-Delgado, S. Peiró, R. Tabarés-Seisdedos, and B. Hutton (2016). "The quality of reporting methods and results of cost-effectiveness analyses in Spain: a methodological systematic review." In: *Systematic Reviews* 5.1, pp. 1–11. ISSN: 2046-4053. URL: `http://dx.doi.org/10.1186/s13643-015-0181-5` (cit. on pp. 34, 104).

Chancellor, J. V., A. M. Hill, C. A. Sabin, K. N. Simpson, and M. Youle (1997). "Modelling the cost effectiveness of lamivudine/zidovudine combination therapy in HIV infection." In: *Pharmacoeconomics* 12, pp. 54–66 (cit. on pp. 26, 49, 70).

Chen, D. S., D. M. Clarrett, L. Li, S. P. Bowditch, J. K. Niparko, and F. R. Lin (2013). "Cochlear implantation in older adults: long-term analysis of complications and device survival in a consecutive series." In: *Otology & Neurology* 34, pp. 1272–1277 (cit. on p. 109).

Chen, J. M., H. Amoodi, and N. Mittmann (2014). "Cost-utility analysis of bilateral cochlear implantation in adults: A health economic assessment from the perspec-

tive of a publicly funded program." In: *Laryngoscope* 124, pp. 1452–1458 (cit. on pp. 93, 94, 104, 111).

Chen, L.-S., C.-S. Liao, S.-H. Chang, H.-C. Lai, and T. H.-H. Chen (2007). "Cost-effectiveness analysis for determining optimal cut-off of immunochemical faecal occult blood test for population-based colorectal cancer screening (KCIS 16)." In: *Journal of medical screening* 14, pp. 191–199 (cit. on p. 125).

Chhatwal, J., S. Jayasuriya, and E. H. Elbasha (2016). "Changing cycle lengths in state-transition models: Challenges and solutions." In: *Medical Decision Making* 36, pp. 952–964 (cit. on pp. 50, 62).

Chouard, C. H., P. Mac Leod, B. Meyer, and P. Pialoux (1977). "Surgically implanted electronic apparatus for the rehabilitation of total deafness and deaf-mutism." In: *Annales d'oto-laryngologie et de chirurgie cervico faciale: bulletin de la Societe d'oto-laryngologie des hopitaux de Paris*. Vol. 94, pp. 353–363 (cit. on p. 90).

Clark, G. (2003). *Cochlear Implants: Fundamentals and Applications*. Springer. ISBN: 978-0-387-95583-4 (cit. on p. 89).

Claxton, K., M. Sculpher, C. McCabe, A. Briggs, R. Akehurst, M. Buxton, J. Brazier, and T. O'Hagan (2005). "Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra." In: *Health Economics* 14, pp. 339–347 (cit. on p. 39).

Corral, J., J. M. Borràs, P. Chiarello, E. García-Alzorriz, F. Macià, A. Reig, J. Mateu de Antonio, X. Castells, and F. Cots (2015). "Estimación del coste hospitalario del cáncer colorrectal en Cataluña." In: *Gaceta Sanitaria* 29, pp. 437–444 (cit. on p. 121).

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag (cit. on p. 24).

Crathorne, L., M. Bond, C. Cooper, J. Elston, G. Weiner, R. Taylor, and K. Stein (2012). "A systematic review of the effectiveness and cost-effectiveness of bilateral multichannel cochlear implants in adults with severe-to-profound hearing loss." English. In: *Clinical Otolaryngology* 37, pp. 342–354 (cit. on p. 91).

de Cock, E., M. Miravitlles, J. R. González-Juanatey, and J. R. Azanza-Perea (2007). "Valor umbral del coste por año de vida ganado para recomendar la adopción de tecnologías sanitarias en España: evidencias procedentes de una revisión de la literatura." In: *Pharmacoeconomics Spanish Research Articles* 4, pp. 97–107 (cit. on pp. 34, 80).

de Dombal, F. T., J. R. Leaper, J. R. Staniland, A. McCann, and J. Horrocks (1972). "Computer-aided diagnosis of acute abdominal pain." In: *British Medical Journal* 2, pp. 9–13 (cit. on p. 13).

Dechter, R. (1996). "Bucket Elimination: A Unifying Framework for Probabilistic Inference." In: *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96)*. Ed. by E. Horvitz and F. V. Jensen. San Francisco, CA: Morgan Kaufmann, pp. 211–219 (cit. on p. 24).

Díez, F. J., M. Luque, and I. Bermejo (2018a). "Decision analysis networks." In: *International Journal of Approximate Reasoning* 96, pp. 1–17 (cit. on pp. 5, 28, 30, 79, 83).

Díez, F. J., M. Luque, J. Pérez-Martín, and M. Arias (2018b). "Research on medical decision analysis at the CISIAD, UNED." In: *Proceedings of the XVIII Conference of the Spanish Association for Artificial Intelligence (CAEPIA-2018)*. Granada, Spain (cit. on p. 133).

Díez, F. J., I. París, J. Pérez-Martín, and M. Arias (2018c). "Teaching Bayesian networks with OpenMarkov." In: *Proceedings of the Ninth European Workshop on Probabilistic Graphical Models (PGM'18)*. Prague, Czech Republic (cit. on pp. 133, 149).

Díez, F. J., M. Yebra, I. Bermejo, M. A. Palacios-Alonso, M. Arias, M. Luque, and J. Pérez-Martín (2017). "Markov influence diagrams: A graphical tool for cost-effectiveness analysis." In: *Medical Decision Making* 37, pp. 183–195 (cit. on pp. 5, 25, 50, 64, 116, 132, 134).

Drummond, M. F., M. J. Sculpher, K. Claxton, G. L. Stoddart, and G. W. Torrance (2015). *Methods for the Economic Evaluation of Health Care Programmes*. cuarta. Oxford, UK: Oxford University Press (cit. on p. 33).

Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddart (2005). *Methods for the Economic Evaluation of Health Care Programmes*. third. Oxford, UK: Oxford University Press (cit. on pp. 32, 104).

Edge, S. B. and C. C. Compton (2010). *AJCC cancer staging manual. 7th edition*. New York: Springer, pp. 1471–1474. ISBN: 978-0-387-88440-0 (cit. on p. 114).

Elbasha, E. H. and J. Chhatwal (2016a). "Myths and misconceptions of within-cycle correction: a guide for modelers and decision makers." In: *Pharmacoeconomics* 34, pp. 13–22. URL: https://doi.org/10.1007/s40273-015-0337-0 (cit. on pp. 49, 57, 60, 62).

Elbasha, E. H. and J. Chhatwal (2016b). "Theoretical foundations and practical applications of within-cycle correction methods." In: *Medical Decision Making* 36, pp. 115–131. URL: https://doi.org/10.1177/0272989X15585121 (cit. on pp. 49, 50, 57, 60, 62).

Estrada, M. D., D. Benítez, P. Clarós, M. A. Clavería, C. Orús, and M. C. Pujol (2011). *Evaluación de los implantes cocleares bilaterales en niños. Criterios de indicación de los implantes cocleares en niños y adultos*. Agència d'Informació, Avaluació i Qualitat en Salut. Informes de Evaluación de Tecnologías Sanitarias, AIAQS núm.2010/03. Madrid, Spain: Ministerio de Ciencia e Innovación (cit. on p. 111).

Euler, L. (1741). "Solutio problematis ad geometriam situs pertinentis." In: *Commentarii academiae scientiarum Petropolitanae*, pp. 128–140 (cit. on p. 12).

Faivre, J., V. Dancourt, C. Lejeune, M. A. Tazi, J. Lamour, D. Gerard, F. Dassonville, and C. Bonithon-Kopp (2004). "Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study." In: *Gastroenterology*. 126, pp. 1674–1680 (cit. on pp. 115, 125).

FDA (2002). *Cochlear implant recipients may be at greater risk for meningitis*. FDA Public Health Web Notification. Available at www.tsbvi.edu/seehear/winter03/fda.htm (cit. on p. 110).

Foteff, C., S. Kennedy, A. H. Milton, M. Deger, F. Payk, and G. Sanderson (2016a). "Cost-utility analysis of cochlear implantation in Australian adults." In: *Otology & Neurotology* 37 (cit. on p. 111).

Foteff, C., S. Kennedy, A. H. Milton, M. Deger, F. Payk, and G. Sanderson (2016b). "Economic evaluation of treatments for pediatric bilateral severe to profound sensorineural hearing loss: An Australian perspective." In: *Otology & Neurotology* 37 (cit. on pp. 108, 111).

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: developed with special reference to the rational foundation of thermodynamics*. Dover Publications (cit. on p. 12).

Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein (1996). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press (cit. on pp. 104, 107).

Gómez Hernández, C., R. Zubizarreta Alberdi, Á. Lanas Arbeloa, V. Valentín Maganto, T. Barata Gomez, F. Perez Riquelme, M. Villacorta González, N. Ascunce Elizaga, C. Cabezas Peña, I. Portillo Villares, M. Prieto García, D. Salas Trejo, R. López García, J. A. Espinàs Piñol, F. Sáez Martínez, E. Almenar Cubells, M. Marzo Castillejo, and J. González Enríquez (2013). *Informe del grupo de expertos sobre concreción de cartera común de servicios para cribado de cáncer*. Tech. rep. Grupo de trabajo sobre concreción de cartera común de servicios del Sistema Nacional de Salud. URL: `http://sectorbarbastro.salud.aragon.es/docs/SECTOR_BARBASTRO_2007.pdf` (cit. on p. 115).

Gondal, G., T. Grotmol, B. Hofstad, M. Bretthauer, T. J. Eide, and G. Hoff (2003). "The Norwegian Colorectal Cancer Prevention (NORCCAP) screening study: baseline findings and implementations for clinical work-up in age groups 50-64 years." In: *Scandinavian journal of gastroenterology* 38, pp. 635–642 (cit. on p. 120).

González, L. P., W. Pignaton, P. S. Kusano, N. S. P. Modolo, J. R. C. Braz, and L. G. Braz (2012). "Anesthesia-related mortality in pediatric patients: A systematic review." In: *Clinics* 67, pp. 381–387 (cit. on pp. 99, 100).

Gorry, G. A. (1973). "Computer-assisted clinical decision making." In: *Methods of Information in Medicine* 12, pp. 45–51 (cit. on p. 13).

Gray, A. M., P. M. Clarke, J. Wolstenholme, and S. Wordsworth (2011). *Applied Methods of Cost-effectiveness Analysis in Healthcare*. New York: Oxford University Press (cit. on pp. 58, 64).

Hamza, S., V. Dancourt, C. Lejeune, J. M. Bidan, C. Lepage, and J. Faivre (2013). "Diagnostic yield of a one sample immunochemical test at different cut-off values in an organised screening programme for colorectal cancer." In: *European Journal of Cancer* 49, pp. 2727–2733 (cit. on p. 125).

Hardcastle, J. D., J. O. Chamberlain, M. H. E. Robinson, S. M. Moss, S. S. Amar, T. W. Balfour, P. D. James, and C. M. Mangham (1996). "Randomised controlled trial of faecal-occult-blood screening for colorectal cancer." In: *The Lancet* 348, pp. 1472–1477 (cit. on p. 115).

Härkönen, K., I. Kivekäs, M. Rautiainen, V. Kotti, V. Sivonen, and J.-P. Vasama (2015). "Sequential bilateral cochlear implantation improves working performance, quality of life, and quality of hearing." In: *Acta oto-laryngologica* 135, pp. 440–446 (cit. on p. 95).

Heckerman, D. E., E. J. Horvitz, and B. N. Nathwani (1992). "Toward normative expert systems: Part I — The Pathfinder Project." In: *Methods of Information in Medicine* 31, pp. 90–105 (cit. on p. 13).

Herranz Amo, F., D. Subirá Ríos, C. Hernández Fernández, J. I. Martínez Salamanca, J. I. Monzó, and R. Cabello Benavente (2006). "Comparación de costes entre la nefrectomía radical laparoscópica y la abierta en el adenocarcinoma renal." es. In: *Actas Urológicas Españolas* 30, pp. 921–925. ISSN: 0210-4806 (cit. on pp. 101, 103).

Hildrew, D. M. and T. B. Molony (2013). "Nucleus N5 CI500 series implant recall." In: *Laryngoscope* 123, pp. 2829–2833. ISSN: 1531-4995 (cit. on p. 99).

Holland, J. F., K. L. Galvin, and R. J. S. Briggs (2012). "Planned simultaneous bilateral cochlear implant operations: How often do children receive only one implant?" In: *International journal of pediatric otorhinolaryngology* 76, pp. 396–399 (cit. on p. 98).

Howard, R. A. and J. E. Matheson (1984a). "Influence diagrams." In: *Readings on the Principles and Applications of Decision Analysis*. Ed. by R. A. Howard and J. E. Matheson. Menlo Park, CA: Strategic Decisions Group, pp. 719–762 (cit. on p. 20).

Howard, R. A. and J. E. Matheson (1984b). *Readings on the Principles and Applications of Decision Analysis*. Menlo Park, CA: Strategic Decisions Group (cit. on p. 29).

Husereau, D., M. Drummond, S. Petrou, C. Carswell, D. Moher, D. Greenberg, F. Augustovski, A. H. Briggs, J. Mauskopf, and E. Loder (2013). "Consolidated Health Economic Evaluation Reporting Standards (CHEERS)–Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force." In: *Value in Health* 16, pp. 231–250 (cit. on p. 39).

International Agency for Research on Cancer (2012). *GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012*. Tech. rep. International Agency for Research on Cancer (cit. on p. 114).

Jensen, F. V. and T. D. Nielsen (2007). *Bayesian Networks and Decision Graphs*. second. New York: Springer-Verlag (cit. on p. 64).

Jensen, F. V., T. D. Nielsen, and P. P. Shenoy (2006). "Sequential influence diagrams: A unified asymmetry framework." In: *International Journal of Approximate Reasoning* 42, pp. 101–118 (cit. on p. 28).

Jensen, F. V. and M. Vomlelová (2002). "Unconstrained influence diagrams." In: *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*

(UAI'02). Ed. by A. Darwiche and N. Friedman. San Francisco, CA: Morgan Kaufmann, pp. 234–241 (cit. on p. 28).

Jensen, F., F. V. Jensen, and S. L. Dittmer (1994). "From influence diagrams to junction trees." In: *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI'94)*. Ed. by R. L. de Mántaras and D. Poole. San Francisco, CA: Morgan Kaufmann, pp. 367–373 (cit. on p. 24).

Johnston, J. C., A. Durieux-Smith, D. Angus, A. O'Connor, and E. Fitzpatrick (2009). "Bilateral paediatric cochlear implants: A critical review." English. In: *International Journal of Audiology* 48, pp. 601–617 (cit. on p. 91).

Kaplan, R. M. and J. W. Bush (1982). "Health-related quality of life measurement for evaluation research and policy analysis." In: *Health psychology* 1, p. 61 (cit. on p. 34).

Kirchhoff, G. (1847). "Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird." In: *Annalen der Physik* 148, pp. 497–508 (cit. on p. 12).

Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: The MIT Press (cit. on pp. 15, 22, 29).

Kolmogorov, A. N. (1933). *Kolmogorov. Foundations of the Theory of Probability*. Chelsea Publishing Company, New York (cit. on p. 12).

König, D. (1936). *Theorie der endlichen und unendlichen Graphen*. Vol. 16. Akademische Verlagsgesellschaft mbh (cit. on p. 12).

Kronborg, O., C. Fenger, J. Olsen, O. D. Jørgensen, and O. Søndergaard (1996). "Randomised study of screening for colorectal cancer with faecal-occult-blood test." In: *The Lancet* 348, pp. 1467–1471 (cit. on p. 115).

Kuntz, K. M. and M. C. Weinstein (2001). "Modelling in economic evaluation." In: *Economic Evaluation in Health Care*. Ed. by M. F. Drummond and A. McGuire. New York: Oxford University Press. Chap. 7, pp. 141–171 (cit. on pp. 64, 67, 74, 75, 80).

Kuthubutheen, J., N. Mittmann, H. Amoodi, W. Qian, and J. M. Chen (2015). "The effect of different utility measures on the cost-effectiveness of bilateral cochlear implantation." In: *Laryngoscope* 125, pp. 442–447 (cit. on pp. 94, 109, 111).

L-Pedraza Gómez, M. J., C. Llorente Parrado, and D. Callejo Velasco (2007). *Implante coclear bilateral en niños: efectividad, seguridad y costes. Situación en las comunidades autónomas*. Madrid: Agencia Laín Entralgo (cit. on pp. 92, 94, 104, 111).

Lacave, C., M. Luque, and F. J. Díez (2007). "Explanation of Bayesian networks and influence diagrams in Elvira." In: *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics* 37, pp. 952–965 (cit. on p. 134).

Lalana, M., A. Fontán, F. J. Díez, J. Pérez-Martín, A. Tapia, J. Millastre, M. M. Larrea, and M. Sánchez (2016a). "Estudio coste-efectividad del cribado de cáncer colorrectal en nuestro sector durante los años 2014-2015." In: *X Congreso Nacional del Laboratorio Clínico*. Zaragoza, España (cit. on pp. 114, 133).

Lalana, M., J. Pérez-Martín, A. Fontán, and F. J. Díez (2016b). "Cost-effectiveness of colorectal cancer screening in Aragon (Spain)." In: *16th Biennial European Conference of the Society for Medical Decision Making*. London, UK (cit. on pp. 114, 132).

Lalwani, A. K. and N. L. Cohen (2011). "Longitudinal risk of meningitis after cochlear implantation associated with the use of the positioner." In: *Otology & Neurotology* 32, pp. 1082–1085 (cit. on p. 109).

Lauritzen, S. L. and D. J. Spiegelhalter (1988). "Local computations with probabilities on graphical structures and their application to expert systems." In: *Journal of the Royal Statistical Society, Series B* 50, pp. 157–224 (cit. on pp. 13, 16).

Lejeune, C., K. Le Gleut, V. Cottet, C. Galimard, G. Durand, V. Dancourt, and J. Faivre (2014). "The cost-effectiveness of immunochemical tests for colorectal cancer screening." In: *Digestive and Liver Disease* 46, pp. 76–81 (cit. on pp. 124, 125).

Levin, B., D. A. Lieberman, B. McFarland, R. A. Smith, D. Brooks, K. S. Andrews, C. Dash, F. M. Giardiello, S. Glick, T. R. Levin, et al. (2008). "Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology." In: *CA: A Cancer Journal for Clinicians* 58, pp. 130–160 (cit. on pp. 118, 125).

Lodwick, G. S., C. L. Haun, W. E. Smith, R. F. Keller, and E. D. Robertson (1963). "Computer diagnosis of primary bone tumors: A preliminary report." In: *Radiology* 80, pp. 273–275 (cit. on p. 13).

López-Bastida, J., F. Sassi, B. Bellas Beceiro, and L. García Pérez (2010). *Análisis coste-efectividad del cribado del cáncer colorrectal en la población general. Madrid: Plan de Calidad para el SNS del MSC. Informes de Evaluación de Tecnologías Sanitarias: SESCS Nº 2006/23.* Tech. rep. Servicio de Evaluación del Servicio Canario de la Salud (cit. on pp. 120, 121, 124–126).

Lovett, R. E. S. (2010). "Comparisons of Unilateral and Bilateral Cochlear Implantation for Children: Spatial Listening Skills and Quality of Life." PhD thesis. University of York, UK (cit. on pp. 95, 109).

Luque, M., F. J. Díez, and C. Disdier (2016). "Optimal sequence of tests for the mediastinal staging of non-small cell lung cancer." In: *BMC Medical Informatics and Decision Making* 16, pp. 1–14 (cit. on pp. 71, 80, 102).

Lutman, M. E. (2008). *Comments on NICE Appraisal of Cochlear Implants (version 2).* Available at www.nice.org.uk/guidance/ta166/documents (cit. on p. 109).

Macrae, F. A., D. J. B. St. John, P. Caligiore, L. S. Taylor, and J. W. Legge (1982). "Optimal dietary conditions for Hemoccult testing." In: *Gastroenterology* 82, pp. 899–903 (cit. on p. 124).

Martín-López, J. E., C. Beltrán-Calvo, R. Rodríguez-López, and T. Molina-López (2014). "Comparison of the accuracy of CT colonography and colonoscopy in the diagnosis of colorectal cancer." In: *Colorectal Disease* 16, O82–O89 (cit. on p. 120).

Naimark, D. M. J., M. Bott, and M. Krahn (2008). "The half-cycle correction explained: two alternative pedagogical approaches." In: *Medical Decision Making* 28, pp. 706–712 (cit. on pp. 48, 50, 54, 60).

Naimark, D. M. J., N. N. Kabboul, and M. D. Krahn (2013). "The half-cycle correction revisited: redemption of a kludge." In: *Medical Decision Making* 33, pp. 961–970 (cit. on pp. 48–50, 54, 56, 57, 60).

Naimark, D. M. J., N. N. Kabboul, and M. D. Krahn (2014). "Response to "The life table method of half-cycle correction: getting it right"." In: *Medical Decision Making* 34, pp. 283–285 (cit. on pp. 48, 60).

Ness, R. M., A. M. Holmes, R. Klein, and R. Dittus (1999). "Utility valuations for outcome states of colorectal cancer." In: *The American journal of gastroenterology* 94, pp. 1650–1657 (cit. on p. 121).

Neumann, P. J., G. D. Sanders, L. B. Russell, J. E. Siegel, and T. G. Ganiats (2016). *Cost-Effectiveness in Health and Medicine*. 2nd. New York: Oxford University Press (cit. on p. 33).

NICE (2009). *Cochlear implants for children and adults with severe to profound deafness. Technology Appraisal Guidance 166*. London, UK (cit. on pp. 108, 110).

NICE (2013). *Guide to the methods of technology appraisal 2013*. National Institute for Clinical Excellence (cit. on pp. 34, 39).

Nielsen, T. D. and F. V. Jensen (1999). "Welldefined decision scenarios." In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*. Ed. by K. Laskey and H. Prade. San Francisco, CA: Morgan Kaufmann, pp. 502–511 (cit. on pp. 22, 27, 29).

Offeciers, E., C. Morera, J. Müller, A. Huarte, J. Shallop, and L. Cavalle (2005). "International consensus on bilateral cochlear implants and bimodal stimulation." In: *Acta Oto-Laryngologica* 125, p. 918 (cit. on p. 91).

Olmsted, S. M. (1983). "On Representing and Solving Decision Problems." PhD thesis. Stanford University, CA: Dept. Engineering-Economic Systems (cit. on p. 64).

Palà, X. F., X. V. Trullén, A. Estrada, R. Clavería, P. Besora, J. M. C. Cortinas, C. Busqué, and E. F. Sallenta (2003). "Análisis del coste económico del tratamiento laparoscópico de la hernia inguinal recidivada: estudio comparativo con la técnica de Lichtenstein." In: *Cirugía Española* 73, pp. 227–232 (cit. on pp. 101, 103).

Papsin, B. C. and K. A. Gordon (2008). "Bilateral cochlear implants should be the standard for children with bilateral sensorineural deafness." In: *Current Opinion in Otolaryngology & Head and Neck Surgery* 16.1, pp. 69–74 (cit. on pp. 89, 91).

Pauker, S. and J. Wong (2005). "The influence of influence diagrams in medicine." In: *Decision Analysis* 2, pp. 238–244 (cit. on pp. 19, 20).

Pearl, J. (1986). "Fusion, propagation and structuring in belief networks." In: *Artificial Intelligence* 29, pp. 241–288 (cit. on p. 13).

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann (cit. on pp. 13, 15).

Peng, R. D. (2011). "Reproducible research in computational science." In: *Science* 334, pp. 1226–1227 (cit. on p. 132).

Pérez-Martín, J., M. A. Artaso, and F. J. Díez (2016). "Cost-utility analysis of bilateral cochlear implants in children." In: *Binaural Hearing with Hearing Implants: A Tribute to 20 Years of Bilateral Cochlear Implantation*. München, Germany (cit. on pp. 88, 133).

Pérez-Martín, J., M. A. Artaso, and F. J. Díez (2017a). "Cost-effectiveness of pediatric bilateral cochlear implantation in Spain." In: *The Laryngoscope* (cit. on pp. 88, 111, 132).

Pérez-Martín, J., I. Bermejo, and F. J. Díez (2017b). "The problem of discontinuities in the evaluation of Markov models." In: *Annual Meeting of the Society for Medical Decision Making*. Pittsburgh, PA (cit. on pp. 48, 133).

Pérez-Martín, J., I. Bermejo, and F. J. Díez (2019). "Evaluation of Markov models with discontinuities." In: *Medical Decision Making* (cit. on pp. 48, 132).

Pérez-Martín, J. and F. J. Díez (2015). "Cost-effectiveness of pediatric bilateral cochlear implantation in Spain." In: *Cochlear Implants Symposium*. Washington, DC (cit. on p. 88).

Peters, B. R. (2007). "Rationale for bilateral cochlear implantation in children and adults." In: *Otology Japan* 17, pp. 603–613 (cit. on p. 91).

Pinto, J. L. and J. E. Martínez (2005). "Estimación del valor monetario de los años de vida ajustados por la calidad: estimaciones preliminares." In: *Ekonomiaz* 60, pp. 192–209 (cit. on p. 34).

Pinto, J. L. and R. Rodríguez (2001). "¿Cuánto vale la pena gastarse para ganar un año de vida ajustado por la calidad? Un estudio empírico." In: *El Valor Monetario de la Salud*. Ed. by J. Puig, J. L. Pinto, and E. Dalmau. Barcelona: Springer, pp. 52–91 (cit. on p. 34).

Prada, L. (2014). "Optimization of Policy Trees for Inluence Diagrams." MA thesis. Madrid, Spain: Dept. Artificial Intelligence, UNED (cit. on p. 134).

Rădulescu, L., S. Cozma, C. Niemczyk, N. Guevara, I. Gahide, J. Economides, J.-P. Lavieille, R. Meller, J.-P. Bébéar, E. Radafy, et al. (2013). "Multicenter evaluation of Neurelec Digisonic® SP cochlear implant reliability." In: *European Archives of Oto-Rhino-Laryngology* 270, pp. 1507–1512 (cit. on p. 99).

Raeve, L. D. (2007). "Education and rehabilitation of deaf children with bilateral cochlear implants: Some practical issues." In: *Bilateral Cochlear Implants: Are Two Always Better Than One?* Nottingham, UK (cit. on p. 91).

Raiffa, H. (1968). *Decision Analysis. Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley (cit. on pp. 17, 39, 64).

Raine, C. H., L. Craddock, and M. E. Lutman (2010). "UK appraisal of clinical and cost-effectiveness of cochlear implantation by the National Institute for Health and Clinical Excellence." English. In: *Cochlear Implants International* 11 Suppl 1, p. 138 (cit. on p. 92).

Ramsden, J. D., B. C. Papsin, R. Leung, A. James, and K. A. Gordon (2009). "Bilateral simultaneous cochlear implantation in children: our first 50 cases." In: *Laryngoscope* 119, pp. 2444–2448 (cit. on p. 98).

Repko, A. F. and R. Szostak (2016). *Interdisciplinary research: process and theory*. SAGE (cit. on p. 132).

Sacristán, J. A., J. Oliva, J. del Llano, L. Prieto, and J. L. Pinto (2002). "¿Qué es una tecnología sanitaria eficiente en España?" In: *Gaceta Sanitaria* 16, pp. 334–343 (cit. on pp. 34, 80).

Sataloff, J. and R. T. Sataloff (2005). *Hearing loss*. Fourth. CRC Press (cit. on p. 88).

Schoonhoven, J. van, M. Sparreboom, B. G. A. van Zanten, R. J. P. M. Scholten, E. A. M. Mylanus, W. A. Dreschler, W. Grolman, and B. Maat (2013). "The effectiveness of bilateral cochlear implants for severe-to-profound deafness in adults: A systematic review." In: *Otology & Neurotology* 34, pp. 190–198 (cit. on p. 91).

Semenov, Y. R., S. T. Yeh, M. Seshamani, N. Y. Wang, E. A. Tobey, L. S. Eisenberg, A. L. Quittner, K. D. Frick, J. K. Niparko, and C. I. Team (2013). "Age-Dependent Cost-Utility of Pediatric Cochlear Implantation." English. In: *Ear and hearing* 34, pp. 402–412 (cit. on p. 108).

Shachter, R. D. (1986). "Evaluating influence diagrams." In: *Operations Research* 34, pp. 871–882 (cit. on pp. 24, 64).

Siebert, U. (2003). "When should decision-analytic modeling be used in the economic evaluation of health care?" In: *European Journal of Health Economics* 4.3, pp. 143–150. ISSN: 1618-7601. URL: `http://dx.doi.org/10.1007/s10198-003-0205-2` (cit. on p. 32).

Siebert, U., O. Alagoz, A. M. Bayoumi, B. Jahn, D. K. Owens, D. J. Cohen, and K. M. Kuntz (2012). "State-transition modeling: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3." In: *Medical Decision Making* 32.5, pp. 690–700 (cit. on pp. 48, 64).

Smulders, Y. E., A. van Zon, G. A. Stegeman I.and van Zanten, A. B. Rinia, et al. (2016). "Cost-utility of bilateral versus unilateral cochlear implantation in adults: A randomized controlled trial." In: *Otology & Neurotology* 37, pp. 38–45 (cit. on pp. 101, 103, 109, 111).

Soares, M. O. and L. Canto e Castro (2012). "Continuous time simulation and discretized models for cost-effectiveness analysis." In: *Pharmacoeconomics* 30, pp. 1101–1117 (cit. on pp. 49, 50, 62).

Sobhani, I., K. Alzahouri, I. Ghout, D. J. Charles, and I. Durand-Zaleski (2011). "Cost-effectiveness of mass screening for colorectal cancer: choice of fecal occult blood test and screening strategy." In: *Diseases of the Colon & Rectum* 54, pp. 876–886 (cit. on pp. 116, 121, 125, 126).

Sociedad Española de Oncología Médica (2017). *Las Cifras del Cáncer en España*. Tech. rep. Sociedad Española de Oncología Médica. URL: `https://seom.org/seomcms/images/stories/recursos/Las_cifras_del_cancer_en_Esp_2017.pdf` (cit. on p. 114).

Sonnenberg, F. A. and J. R. Beck (1993). "Markov models in medical decision making: A practical guide." In: *Medical Decision Making* 13, pp. 322–339 (cit. on pp. 25, 48, 54, 60, 64).

Soto, J. (2004). "Nuevos medicamentos en España: ¿Cuándo deberían ser considerados alternativas coste-efectivas e inversiones rentables para el Sistema Nacional de Salud?" In: *Farmacia Hospitalaria* 28, pp. 296–304 (cit. on p. 34).

Sparreboom, M., J. van Schoonhoven, B. G. A. van Zanten, R. J. P. M. Scholten, E. A. M. Mylanus, W. Grolman, and B. Maat (2010). "The effectiveness of bilateral cochlear implants for severe-to-profound deafness in children: A systematic review." English. In: *Otology & Neurotology* 31, pp. 1062–1071 (cit. on p. 91).

Sparreboom, M., A. F. M. Snik, and E. A. M. Mylanus (2012). "Sequential bilateral cochlear implantation in children. Quality of life." English. In: *Archives of Otolaryngology–Head & Neck Surgery* 138, pp. 134–141 (cit. on pp. 95, 109).

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley, New York (cit. on p. 35).

Stinnett, A. A. and J. Mullahy (1998). "Net health benefit: A new framework for the analysis of uncertainty in cost-effectiveness analysis." In: *Medical Decision Making* 18, S68–80 (cit. on p. 34).

Sucar, L. E. (2015). *Probabilistic Graphical Models: Principles and Applications*. Springer. ISBN: 978-1-4471-6698-6 (cit. on pp. 15, 39).

Summerfield, A. Q., R. E. S. Lovett, H. Bellenger, and G. Batten (2010). "Estimates of the cost-effectiveness of pediatric bilateral cochlear implantation." English. In: *Ear & Hearing* 31, pp. 611–624 (cit. on pp. 93–95, 98–103, 108, 109, 111).

Summerfield, A. Q., D. H. Marshall, G. R. Barton, and K. E. Bloor (2002). "A cost-utility scenario analysis of bilateral cochlear implantation." English. In: *Archives of Otolaryngology–Head & Neck Surgery* 128, pp. 1255–1262 (cit. on pp. 92–95, 109).

Summerfield, A. Q., J. Osborne, M. Doran, R. Ramsden, D. Mawman, M. O'Driscoll, J. Graham, W. Aleksy, L. Meerton, C. Verschure, P. Ashcroft, G. R. Barton, M. Pringle, J. Toner, C. McAnallen, D. Proops, C. Harries, H. Cooper, I. Court, and R. Gray (2006). "Self-reported benefits from successive bilateral cochlear implantation in post-lingually deafened adults: randomised controlled trial." English. In: *International Journal of Audiology* 45, S99–S107 (cit. on pp. 94, 95).

Summerfield, A. Q., P. C. Stacey, K. L. Roberts, H. M. Fortnum, and G. R. Barton (2003). "Economic analysis and cochlear implantation." English. In: *International Congress Series* 1254, pp. 313–319 (cit. on pp. 91–94).

Szolovits, P. and S. G. Pauker (1978). "Categorical and probabilistic reasoning in medicine." In: *Artificial Intelligence* 11, pp. 115–144 (cit. on p. 13).

Tan-Torres Edejer, T., R. Baltussen, T. Adam, R. Hutubessy, A. Acharya, D. B. Evans, and C. J. L. Murray, eds. (2012). *Making Choices in Health: WHO Guide to Cost-Effectiveness Analysis*. Geneva, Switzerland: World Health Organization (cit. on p. 39).

Tappenden, P., J. Chilcott, S. Eggington, H. Sakai, J. Karnon, and J. Patnick (2007). "Option appraisal of population-based colorectal cancer screening programmes in England." In: *Gut.* 56, pp. 677–684 (cit. on pp. 116, 119, 120, 125, 126).

Tatman, J. A. and R. D. Shachter (1990). "Dynamic programming and influence diagrams." In: *IEEE Transactions on Systems, Man and Cybernetics* 20, pp. 365–379 (cit. on p. 27).

Torre, A. de la, M. Sainz, J. L. Padilla, C. Roldán, J. García-Valdecasas, R. M. Contreras, C. Maroto, et al. (2005). *Evaluación de la efectividad terapéutica del tratamiento de hipoacusias severas y profundas en niños: implante coclear versus audífono. Costes, beneficios y mejora en la calidad de vida.* Memoria final del proyecto 03/10010 (cit. on pp. 101, 103).

Tosh, J. and A. Wailoo (2008). *Review of software for decision modelling.* Tech. rep. Health Economics, Decision Science, School of Health, and Related Research, University of Sheffield (cit. on p. 65).

Towler, B., L. Irwig, P. Glasziou, J. Kewenter, D. Weller, and C. Silagy (1998). "A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, hemoccult." In: *BMJ: British Medical Journal* 317, pp. 559–565 (cit. on p. 115).

Turing, A. M. (1950). "Computing Machinery and Intelligence." In: *Mind* 59, pp. 433–460 (cit. on p. 12).

Tutte, W. T. (2001). *Graph theory.* Cambridge: Cambridge University Press. ISBN: 978-0-521-79489-3 (cit. on p. 12).

Ubel, P. A., R. A. Hirth, M. E. Chernew, and A. M. Fendrick (2003). "What is the price of life and why doesn't it increase at the rate of inflation?" In: *Archives of internal medicine* 163, pp. 1637–1641 (cit. on p. 34).

van der Gaag, L. C. and V. M. H. Coupe (2000). "Sensitivity analysis for threshold decision making with Bayesian belief networks." In: *Lecture Notes in Artificial Intelligence* 1792, pp. 37–48 (cit. on p. 41).

Versteegh, M., S. Knies, and W. Brouwer (2016). "From good to better: new Dutch guidelines for economic evaluations in healthcare." In: *PharmacoEconomics*, pp. 1071–4 (cit. on p. 39).

Walker, S., F. Girardin, C. McKenna, S. Ball, J. Nixon, S. Plein, J. P. Greenwood, and M. Sculpher (2013). "Cost-effectiveness of cardiovascular magnetic resonance in the diagnosis of coronary heart disease: an economic evaluation using data from the CE-MARC study." In: *Heart* 99, pp. 873–881 (cit. on p. 80).

Wang, J. T., A. Y. Wang, C. Psarros, and M. da Cruz (2014). "Rates of revision and device failure in cochlear implant surgery: A 30-year experience." English. In: *Laryngoscope* 124, pp. 2393–2399 (cit. on pp. 99, 100).

Warner, H. R., A. F. Toronto, and L. G. Veasy (1964). "Experience with Bayes' theorem for computer diagnosis of congenital heart disease." In: *Annals of the New York Academy of Sciences* 115, pp. 558–567 (cit. on p. 13).

Weinstein, M. C., H. V. Fineberg, A. S. Elstein, et al. (1980). *Clinical Decision Analysis*. Philadelphia, PA: Saunders (cit. on p. 144).

Weinstein, M. C., B. O'Brien, J. Hornberger, J. Jackson, M. Johannesson, C. McCabe, and B. R. Luce (2003). "Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices-Modeling Studies." In: *Value in Health* 6, pp. 9–17 (cit. on p. 39).

Weinstein, M. C., J. E. Siegel, M. R. Gold, M. S. Kamlet, and L. B. Russell (1996). "Recommendations of the panel on cost-effectiveness in health and medicine." In: *JAMA*. 276, pp. 1253–1258 (cit. on p. 122).

Winawer, S. J., A. G. Zauber, M. J. O'brien, M. N. Ho, L. Gottlieb, S. S. Sternberg, J. D. Waye, J. Bond, M. Schapiro, E. T. Stewart, et al. (1993). "Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps." In: *New England Journal of Medicine* 328, pp. 901–906 (cit. on p. 120).

Wold, H. (1954). "Causality and econometrics." In: *Econometrica: Journal of the Econometric Society*, pp. 162–177 (cit. on p. 13).

Wright, S. (1921). "Correlation and causation." In: *Journal of Agricultural Research* 20, pp. 557–585 (cit. on p. 12).