Multi-Criteria Decision Analysis (MCDA) as the basis for the development, implementation and evaluation of interactive patient decision aids

FRANCISCO POZO-MARTIN

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

November 2015

Department of Social and Environmental Health Research
Faculty of Public Health and Policy
LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE,
UNIVERSITY OF LONDON

Funding details – no funding received

I, Francisco Pozo-Martin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: …….

ABSTRACT

BACKGROUND: In the context of the progressive movement towards patient-centred care, patient-specific decision support is an important focus of interest. Many diagnostic and treatment patient decision aids (PDAs) are now available to help patients make informed choice decisions. An increasing number of these are software-based, with some available online. Multi-Criteria Decision Analysis (MCDA) is a potentially useful technique on which to base a software-assisted PDA, especially when the decision is complex - as is the case in choosing the best treatment for non-small cell lung cancer – but it has so far been relatively little exploited in this area. The use of any from a number of existing MCDA-based software applications in the development and delivery of a MCDA-based interactive PDA can be an effective way of achieving "best-practice" or normative standards of decision making, such as 1) a well-constructed set of decision criteria or 2) logically consistent patient preferences. However, it also involves the use of resources such as the time and cognitive effort involved in decision-making. The comparative evaluation of alternative MCDA-based software applications in developing and delivering a PDA therefore involves trade-offs between decision effectiveness and decision resource criteria moving from the normative to the prescriptive. MCDA is an ideal tool for this meta-evaluation task as well as for the adoption decision itself.

AIM: To analyse, as proof of concept, the use of MCDA for the development, implementation and evaluation of interactive PDAs in routine clinical practice.

OBJECTIVES:

1. To assess the use with clinicians in the Spanish NHS of two alternative MCDA software applications which implement dissimilar MCDA techniques in the development of a PDA in routine clinical practice;

2. To assess the use with clinicians in the Spanish NHS of the same two alternative MCDA software applications in the implementation of a PDA in an environment replicating actual clinical consultations;

3. To build a meta-multi-criteria decision model based on the Decision Resources Decision Effectiveness Analysis (DRDEA) framework and assess the use of this model by clinicians in the Spanish NHS to make the choice between the two MCDA applications as the basis for a PDA.

METHODS:

1) Two dissimilar MCDA software applications served as a basis for the development of a lung cancer clinical management PDA in close collaboration with two different groups of three clinicians from two different Spanish NHS hospitals (H1 and H2): 1) *Expert Choice*, which implements the Analytic Hierarchy Process (AHP) MCDA approach; 2) *Annalisa in Elicia* (*ALEL*), which implements the Simple Attribute Weighting (SAW) MCDA approach. The process of co-development of the PDA in hospitals H1 and H2 was documented;

2) *Expert Choice* was used to implement (i.e. deliver) the lung cancer clinical management PDA in three hypothetical consultations in hospital H1. In each consultation, one of the three clinicians involved in the development of the tool, with support by this researcher, guided a proxy patient (a non-clinical member of hospital staff) through the PDA. The same process was repeated with the MCDA software *ALEL* in hospital H2. The process of delivery of the PDA in hospitals H1 and H2 was documented;

3) This researcher built a meta-multi-criteria decision model based on the DRDEA framework to help clinicians choose between different MCDA software applications as the basis of a PDA. The MCDA approach used for this meta-model was Multi-Attribute Value Theory (MAVT). The model was implemented, using the software HiView 3, with three clinicians from hospital H3 for the choice between *Expert Choice* and *ALEL* as the basis of a lung cancer clinical management PDA.

RESULTS:

The thesis makes a three-fold contribution to research in patient-centred decision support. First, it presents two new MCDA software-based approaches to clinical decision support, based on joint work with clinicians in the Spanish NHS, for developing an interactive PDA for the clinical management of non-small cell lung cancer. Second, it describes the use of these decision support tools in the delivery of

an interactive PDA for the clinical management of non-small cell lung cancer in a hospital environment via simulated consultations between actual clinicians, with support from this researcher, and proxy lung cancer patients. Third, it presents and applies a new MCDA-based methodology for evaluating the use of alternative MCDA software applications in the development and delivery of interactive PDAs.

AKNOWLEDGEMENTS

The conversation that this thesis originated from took place on the street in Johannesburg, South Africa, several years ago. Much has happened since then. Throughout the years that led to this day, Zaid Chalabi and Jack Dowie have been a constant source of knowledge and support. Thank you. As Jack said to me once, quoting Shoeless Joe Jackson, "if you build it they will come", and he was right. Apart from Zaid and Jack, other friends have illuminated the way. Some quite literally, as they are the wonderful clinicians that participated in this research: Angel Salvatierra, Eduard Monsó, Jaume Sauleda, Miguel Angel Martín, Isidoro Barneto, Teresa Morán, Pedro López de Castro, Mónica Guillot, and Miguel Angel Carvajal. In addition to them, my friends Angel López Encuentra and Víctor Abraira provided me with extremely valuable insights at a difficult moment.

Light also came from Hilary, Caroline, Alvaro, Douglas, Charlotte, Angel, Rosa, Luciano, Miguel, Alex, Sara, James, Clara, Yoel, Thanassis, Isabelle, Chris, Sandra, Andy, Melisa, Triantafyllos, Mariana, Catriona, Sotiris, Karima, Alessandra, Ana, Milica, David, Fernando, Leandro, Alejandra, and Yoshimi. Thank you.

Judy, thank you! Karl, merci bien!

Paco and Wigberta, my parents, continue giving to me unconditionally, and so do Gonzalo and María, Daniel, and Ula. Thank you.

This thesis is dedicated to my wife Stephanie, my daughter Clara, and my son Eloy.

Table of Contents.

List of Figures.

List of Tables:

Introduction:

In recent years there has been an increased emphasis in involving patients in making decisions about their own health care [1]. Different models of clinical decision making where the patient is an active participant in the decision making process have been proposed. One model that has gained high policy support is Shared Decision Making (SDM) [2, 3]. SDM is a concept variably and loosely defined [4]. In a review of the literature on alternative definitions of SDM, Makoul and Clayman [4] proposed that an integrative model of SDM would have the following elements: a discussion between the patient and the clinician of the problem to be addressed, the options available, and their pros and cons; an explication of the patient's values and preferences, as well as of the physician's knowledge and recommendations; an assessment of patient self-efficacy to adhere to a plan of action; an explicit decision (or an explicit deferment of a decision) and follow-up [2].

SDM can be facilitated by the use of patient decision aids [5]. Patient decision aids (PDAs) are "evidence-based tools designed to help patients participate in making specific and deliberated choices among health care options" [5]. In general, they provide decision support to patients by 1) making explicit the decision that needs to be made, 2) providing evidence-based information about the condition, the options, the consequences (benefits, harms), their probabilities and the uncertainties related to each of the health care options under consideration, and 3) helping patients express their values/ preferences with regards to the benefits, harms and uncertainties of the options [6]. PDAs are available in different formats [7]. A number of them are available on the internet. They can be self-administered by patients or used with practitioners in the consultation [7].

A systematic review of randomised clinical trials (RCT) of PDA effectiveness provides enough evidence that PDAs improve decision making (in terms e.g. of improving patient knowledge, reducing decisional conflict and increasing the consistency between patient values and the chosen option) so as to warrant their use in clinical practice [5]. However, the routine implementation of these tools to support SDM has challenges [7]. Among the barriers cited in the literature by practitioners are 1) the lack of skills to practice shared decision making, 2) an organisational

culture that hinders the uptake of PDAs, 3) the perception that PDAs are inappropriate and/or too complex for use with certain groups of patients, 4) problems with workflow fit [7]. These barriers highlight the importance of developing "tools, processes and systems of care which make SDM feasible given the constraints of routine clinical practice" [8].

Multi-Criteria Decision Analysis (MCDA) can be the used as the basis for such tools [8]. MCDA is "an umbrella term for a collection of formal approaches which explicitly seek to take into account of multiple criteria in helping individuals or groups explore decisions that matter" [9]. In preference-sensitive choice decision situations such as those which are often the subject of SDM, a number of MCDA techniques are available to provide decision support. For these decisions, the MCDA process usually involves the following steps: 1) Identifying the decision problem, 2) clearly specifying the options, 3) Identifying the objectives of the decision and related measureable criteria, 4) measuring the consequences of the options on each of the criteria and, depending on the specific MCDA technique used, scoring these consequences on a common scale, and 5) assigning weights of relative importance to the criteria [10]. Formal procedures are used to combine the information from steps 4) and 5) to provide a recommendation for the decision, usually in the form of a ranking of the options from most to least preferred. A sixth step in the process involves analysing the robustness of the results to changes in inputs that are not defined or measured precisely. The main role of MCDA is to help decision-makers manage large amounts of complex information in a way that is consistent [10].

Examples of well-known MCDA approaches include Multi-Attribute Value Theory (MAVT) [11], the Analytic Hierarchy Process (AHP) [12], the family of methods Election et Choix Traduisant la Realite (ELECTRE) [13] and the family of methods Preference Ranking Organisation Method for Enriched Evaluation (PROMETHEE) [14]. The implementation of these methods is usually done using decision analytic software applications. For example, HiView 3 [15] implements MAVT, and Expert Choice [16] implements the AHP.

MCDA can and has been used in decision support for complex clinical decisions, such as in oncology [17]. In addition, some MCDA approaches have been shown to be acceptable to patients and within their capabilities [18].

Clinicians can use different software applications for developing and implementing MCDA-based PDAs for complex clinical decisions in their day-to-day clinical practice environment. According to the Decision resource-decision effectiveness analysis (DRDEA) framework [19], the choice by clinicians between alternative software applications or "templates" for developing and implementing MCDA-based PDAs in routine clinical practice is a multi-criterion meta-decision that can be expressed in terms of the trade-offs between two types of criteria: 1) decision resources (DR) criteria expressing the resource requirements associated with using each template (e.g. time required, cognitive effort required, or financial cost [19]) and 2) decision effectiveness (DE) criteria expressing the benefits of using each template (e.g. theoretical grounding, evidential strength and coverage, transparency [19]). This choice depends on the specific context in which clinicians operate and it is preference-sensitive, i.e. it depends on how clinicians trade-off DR and DE criteria. According to DRDEA [19], MCDA is the appropriate technique to make this choice.

Aim: To analyse, as proof of concept, the use of MCDA for the development, implementation and evaluation of interactive patient decision aids in routine clinical practice

Objectives.

1. To assess the use with clinicians in the Spanish NHS of two alternative MCDA software applications which implement dissimilar MCDA techniques in the development of a PDA in routine clinical practice;

2. To assess the use with clinicians in the Spanish NHS of the same two alternative MCDA software applications in the implementation of a PDA in an environment replicating actual clinical consultations

3. To build a meta-multi-criteria decision model based on the DRDEA framework and assess the use of this model by clinicians in the Spanish NHS to make the choice between the two MCDA applications as the basis for a PDA.

The case study.

The case study in this research is a hypothetical 69 year-old patient recently diagnosed with stage $IIIA_3$ non-small cell lung cancer and with lung and

cardiovascular comorbidities who will use an MCDA-based PDA to choose between a set of available clinical management strategies for his condition.

Methods.

1. The two MCDA software applications which served as a basis for the development of the lung cancer PDA were 1) *Expert Choice* [16], which implements the Analytic Hierarchy Process (AHP) MCDA approach, and *Annalisa in Elicia* (ALEL) [20], which implements the Simple Additive Weighting (SAW) MCDA approach.

Expert Choice was used by this researcher to co-develop with 3 clinicians (one oncologist, one pulmonologist, one thoracic surgeon) from hospital H1 in the Spanish NHS a Stage $IIIA_3$ lung cancer clinical management PDA. The same process was repeated with a different group of 3 clinicians working in hospital H2 in the Spanish NHS using *ALEL*.

The process of co-development of each PDA with clinicians was documented by this researcher. This process included the following steps: 1) determining the options, 2) determining the relevant criteria for the decision, 3) measuring the consequences of the options on the criteria, and 4) scoring these consequences.

2. *ALEL* was used to implement (i.e. deliver) the lung cancer clinical management PDA in 3 hypothetical consultations in hospital H1. In each consultation, one of the three clinicians involved in the development of the tool, with support by this researcher, guided a proxy patient (a non-clinical member of hospital staff) through the PDA. The same process was repeated with the MCDA software Expert Choice in hospital H2.

The process of implementation of each PDA in each hypothetical consultation, documented by this researcher, included the following steps: 1) communication of the criteria, 2) preference elicitation, 3) communication of the options, 4) communication of the results of the decision, 5) communication of the evidence, 6) sensitivity analysis. In addition, at the end of each consultation the perceived quality of the decision of both the physician and the patient was assessed using the "My Decision Quality" (MDQ) questionnaire [21].

3. This researcher built a meta-multi-criteria decision model based on the DRDEA framework to help clinicians choose between different MCDA software applications

as the basis of a PDA. The MCDA approach used for this meta-model was Multi-Attribute Value Theory [11]. The model was implemented with 3 clinicians (one oncologist, one pulmonologist, one thoracic surgeon) from hospital H3 in the Spanish NHS for the choice between *ALEL* and Expert Choice as the basis of a lung cancer management PDA. The MCDA software used to implement the meta-model was HiView 3 [15].

The process of development of the meta-model by this researcher had four steps: 1) determining the options, 2) determining the relevant decision effectiveness (DE) and decision resources (DR) criteria, 3) measuring the consequences of the options on the criteria, and 4) scoring these options. The implementation of the meta-model with each of the 3 clinicians from hospital H3 had three steps: 1) preference elicitation for the different DE and DR criteria, 2) review of results, 3) sensitivity analysis.

This research project was considerably informed by the concept of engaged scholarship, which involves generating knowledge in collaboration with practitioners (in this case, clinicians from three hospitals in the Spanish NHS) that can jointly advance the scientific enterprise and enlighten the community of those practitioners [22]. In this sense, the methods used were context-led. That is, they evolved in response to the conditions found in the context of clinical practice where the research was conducted. Particularly, in response to the time constraints posed to the clinicians involved. This should not be considered a limitation but the essence of this study.

Structure of the thesis.

This thesis is structured to analyse, as proof of concept, the use of MCDA for the development, implementation and evaluation of patient decision aids in routine clinical practice. Chapter 1 presents a literature review of interactive PDAs. This includes a review of the current status in the field of PDAs with a focus on successful empirical applications, a review of MCDA methods, and a review of software applications supporting MCDA. The chapter ends with a justification of the case study for this thesis: a hypothetical 69 year-old patient recently diagnosed with stage $IIIA_3$ non-small cell lung cancer and with lung and cardiovascular comorbidities who will use an MCDA-based PDA to choose between a set of available clinical management strategies for his condition. Chapter 2 describes the process and results

of using Expert Choice and *ALEL* with clinicians to develop a Stage $IIIA_3$ lung cancer clinical management PDA in routine clinical practice. Chapter 3 describes the process and results of using Expert Choice and *ALEL* with clinicians to implement a Stage $IIIA_3$ lung cancer clinical management PDA in an environment replicating actual clinical consultations. Chapter 4 describes the process and results of developing and implementing a meta-multi-criteria decision model based on the DRDEA framework to help clinicians choose between Expert Choice and *ALEL* as the basis for a Stage $IIIA_3$ lung cancer clinical management PDA. The chapter shows that the decision is preference-sensitive: it depends on the trade-offs clinicians are willing to make between "decision resources" and "decision effectiveness" criteria. Chapter 5 brings all the elements of the thesis together by presenting a discussion of the main findings and lessons learnt, as well as suggestions for further research.

Chapter 1. Literature review.

1. What is and what is not a PDA.

PDAs are tools to help patients participate in making decisions about health care options [5]. They contain, at least, information about the health care options and about their consequences (e.g. benefits and harms) on an individual's health [6]. In addition, they may provide 1) information on the condition, 2) the chances (i.e. probabilities) that the patient will experience these consequences, 3) an explicit exercise to help patients clarify their preferences (i.e. values) over the consequences of the options, 4) other's testimonials, and 5) assistance through the decision-making process [6].

PDAs are most often used to help patients make preference-sensitive decisions, i.e. decisions for which the choice of option is not clear (e.g. because several options with similar efficacy are available) and thus depend on the patient's valuation of the different consequences of the options [23]. They support the process of Shared Decision Making, a mode of decision making in which both clinicians and patients share information in order to reach a consensus about the preferred treatment [24].

The above definition of PDAs excludes 1) passive informed consent materials, 2) educational interventions which are not targeted to making a specific decision, 3) interventions to increase adherence to a particular option. It also excludes computerised decision support systems (CDSS). CDSS are information technology (IT)-based systems that use algorithms to provide specific recommendations to clinicians about a particular patient or his/her condition [25], including, *inter alia,* computerised physician order entry systems, reminder systems, alert systems, and medical calculators. CDSS are not patient-centred but clinician-centred systems.

2. Origin, growth, drivers.

The early development of PDAs was influenced by work on decision support tools based on Decision Analysis [23], a quantitative approach to decision making first used in patient counselling in the late 1970s [26, 27]. In the 1980s, the work by Mulley, Wennberg and others [28, 29] on interactive multimedia programmes for the treatment on benign prostate hyperplasia was influential in the development of the

field [23]. PDAs are primarily developed and used in Australia, North America and Europe [5]. In the last 15 years their number has proliferated rapidly [5]. While in 1999 only 15 PDAs had been developed by researchers [30], in 2007 O'Connor reported the existence of more than 450 PDAs and that the previous year more than 8 million hits were made on the websites of the main PDA providers [31]. In 2014, the Ottawa Decision Aid Repository, a comprehensive collection of PDAs available on the internet, received more than 80,000 visits and provided information about 647 PDAs [32].

Several factors have influenced the development of PDAs, among which are the following:

1. An acknowledgement that, with the rapid expansion of health care interventions in recent years, several options are often available for a particular clinical decision, and that the choice of option will often depend on the trade-offs between the harms, benefits, and uncertainties related to each of these options [1];

2. The search for interventions to help reduce unwarranted variations in the provision of health care (i.e. variation in the provision of care that could not be explained by the need for these interventions) [23, 33]. In this sense, the use of PDAs is intended to spur patient self-interest in avoiding those interventions [23];

3. The increased importance of the ethical principle of respect for patient autonomy [23], which promotes that patients make choices understanding all the relevant information about the options and which has prompted changes in the legal requirements for informed consent [34]. These changes highlight the importance that patients are fully informed of the consequences of the different health care options available to them before they make a decision [34]. PDAs are tools which may support this process;

4. The influence of the evidence-based medicine (EBM) movement [6]. EBM highlights the importance of making available and taking into account information about the outcomes of different health care interventions in clinical decision making.

3. Types of PDAs.

PDAs are heterogeneous tools. In terms of their scope, they vary in:

- The health care condition they address, ranging from allergies to depression, from cancer to end of life care [32];

- The type of decision they support, such as choice of screening test, choice of treatment, or choice of self-management strategy [1];

In terms of their characteristics, they differ, *inter alia*, in terms of:

- Their format, including booklets (see, for example Labrecque et al. [35] or Legare et al. [36]), audiotape workbooks (e.g. Hunter et al. [37]), videos (e.g. Partin et al. [38] and Pignone et al. [39]), and internet-based applications such as those provided by the following online resources: Patient [40], NHS Rightcare [41], Healthwise [42], Mayo Clinic Shared Decision Making National Resource Center [43];

- Their mode of delivery, either self-administered by patients or used with clinicians in the consultation [7];

- Their components, which may or may not include probabilistic information about the consequences of the options, visual representations of this probabilistic information, explicit value clarification exercises, personal stories (e.g. testimonials of other patients) or guidance through the decision-making process [5];

- The theory or model of decision making which informs their design and development [26], e.g. Decision Analysis under Expected Utility Theory (e.g. Bekker at al. [44] or Montgomery et al. [45]), Multi-Criteria Decision Analysis (e.g. Dolan and Frisina [46]), the Ottawa Decision Support Framework (e.g. Hunter et al. [37] or Lalonde et al. [47]), the Health Belief Model (e.g. Schapira and VanRuiswyk [48]), or no theory or model of decision making (e.g. Auvinen et al. [49] or Deschamps et al. [50]).

For the purposes of this thesis, one type of PDAs is particularly relevant: computerised PDAs. These PDAs use a computerised medium, and their format includes video-discs, CD-ROM or computer-based programmes, and websites [51]. Computerised PDAs have a number of advantages over other types of PDAs: 1) they allow for interactivity and for the use of visual features which can facilitate patient involvement, 2) they can include information tailored to specific patients, 3) they can provide feedback to increase understanding and 4) they can facilitate the dissemination of information [51]. PDAs are increasingly computerised [51].

4. What makes a good PDA? Different views on the most appropriate measure of PDA effectiveness.

In principle a good PDA is a PDA that leads to a good decision. However, there are different views about what constitutes a good decision [5]. These different views are reflected in the literature in a debate around what should be the focus of the evaluation of PDAs, i.e. what is the most appropriate measure of the effectiveness of PDAs [52].

Studies measuring the effectiveness of PDAs have used a wide range of outcomes. The next five sections provide an overview of these outcomes. Incorporated in the narrative are any salient criticisms and justifications in the literature for the use of these outcomes as measures of effectiveness of PDAs.

4.1. Outcomes related with decision-making.

Decision-making using a PDA involves patients engaging in a decision-making process to make a choice. The next two sub-sections explores PDA effectiveness outcomes related with 1) the decision making process and 2) the choice made. A third sub-section section explores a number of constructs of decision quality, which are multidimensional measures of PDA effectiveness.

4.1.1. Outcomes related with the decision-making process.

Outcomes related with the decision-making process include variables such as 1) preparation for decision-making, 2) satisfaction with decision-making, 3) communication between the patient and the practitioner, 4) patient involvement in decision making, and 5) decisional conflict [5].

Preparation for decision-making refers to the patient's perception of the usefulness of a PDA in helping him/her prepare to communicate with the practitioner and to make a decision [53]. It has been measured using e.g. the Preparation for Decision Making Scale [5, 53].

Different measures of satisfaction include self-reported satisfaction with the option chosen and satisfaction with the process of decision-making [5]. These measures have been criticised by some authors as inadequate measures of PDA effectiveness

because satisfaction depends on a patient's expectations [54] rather than in the quality of decision support [55].

Communication between the patient and the practitioner has been measured [5] using instruments such as the Observing Patient Involvement in Decision Making (OPTION) scale [56] or the Informed Decision Making instrument [57].

Patient involvement in decision making has been measured [5] using e.g. the Control Preferences Scale (CPS), which identifies the preferred role of the patient in terms of decisional control (active, shared with the clinician, or passive) [58]. Measures highlighting patient participation in the consultation have been considered by some researchers as unsuitable for primary measures of PDA effectiveness on the grounds that the role of PDAs is not to promote any particular model of decision-making [52].

Decisional conflict is a construct measuring an individual's "level of personal uncertainty about the course of action to take" [59]. Factors influencing decisional conflict in patients include the absence of information about the health care options and their consequences, values which are not clear, insufficient skills in decision-making or in putting decisions into practice, emotional distress, and the perception that significant others exercise pressure to impose their views on the decision [59]. A commonly used measure of decisional conflict is the Decisional Conflict Scale (DCS), an instrument composed of 16 items grouped into five subscales measuring patients' perceptions in terms of the following domains: feeling uncertain in making a choice, feeling uninformed, feeling unclear about values, feeling unsupported in decision making, feeling that the decision was effective [60]. Decisional conflict has been criticised as a measure of PDA effectiveness by some authors on the grounds that focusing on reducing decisional conflict penalises decision processes which create in patients a state of warranted equipoise with regards to the decision [21].

4.1.2. Outcomes related with the choice made.

Outcomes related with the choice made include 1) knowledge acquired by the patient about the condition, the options and their consequences and 2) accuracy in the patient's perceived probabilities of the options' consequences and 3) concordance between the option chosen and the patient's values [5].

Knowledge is measured using knowledge questionnaires, generally based on multiple choice or true/false questions and usually designed specifically for each study [61]. Perceived probabilities of the options' consequences are measured using e.g. probability scales [62, 63], which are then matched to the actual probabilities presented in the PDAs to assess their accuracy. Some authors have argued against knowledge and accuracy in perceived probabilities, two types of knowledge outcomes [52], as being the most appropriate measures of PDA effectiveness [36, 64]. Knowledge makes patients more informed and thus helps them deliberate about their preferences, but it does not ensure, by itself, that patients will be more involved in decision making or that they will get the option that they desire [36]. In addition, knowledge as a key element of a good decision is not a concept easy to operationalise [64]. For example, does there exist an amount of knowledge which is both necessary and sufficient for each decision question or should patients determine when they know enough? [64].

Achieving concordance between the option chosen and the patient's values has been described as one important measure of PDA effectiveness [52] and increasingly identified as a key goal of PDAs [65], in line with the argument that PDAs should primarily help patients make personalised choices between options [52]. Studies have measured value-choice concordance using different approaches, e.g. calculating the percentage of patients getting a treatment that matched their stated preferences or eliciting patients' preferences and using regression models to predict patients choices based on those preferences [5].

4.1.3. Constructs of decision quality

A number of constructs of decision quality have been proposed in the literature to measure the effectiveness of PDAs. Perhaps the most well-known of these constructs is the one developed by the International Patient Decision Aids (IPDAS) Collaboration. The IPDAS Collaboration, founded in 2003, is an international group of researchers, PDA developers and other stakeholders with the main task of developing a framework to improve the content, development, implementation and evaluation of PDAs [66]. The IPDAS Collaboration has developed a consensus definition of PDA effectiveness based on two main components: 1) the quality of the decision-making process and 2) the quality of the choice made [30]. According to

this definition, a PDA fosters a high quality decision process if it helps patients "recognise that a decision needs to be made", "feel informed about the options and their features", "understand that values affect the decision", "be clear about the option-features that matter most", "discuss values with their practitioner", and "become involved in their preferred ways" [67]. A PDA fosters a high quality choice if it "improves the match between the chosen option and the patients' values" [67]. The IPDAS definition of PDA effectiveness combines different outcomes of the decision making process and of the choice made explored in previous sub-sections. Table 1.1 shows the different elements of the IPDAS definition of PDA effectiveness against the outcomes that map onto these elements [5, 53, 61].

Table 1.1. IPDAS definition of PDA effectiveness.

Elements of the definition	**Outcomes of PDA effectiveness mapping onto the elements of the definition**
Quality of the decision-making process:	
1. "Recognise that a decision needs to be made"	1. Preparation for decision-making
2. "Feel informed about the options and their features"	2. Decisional conflict (feeling uninformed)
3. "Understand that values affect the decision"	3. Preparation for decision-making
4. "Be clear about the option-features that matter most"	4. Decisional conflict (feeling unclear about values)
5. "Discuss values with their practitioner"	5. Patient-practitioner communication
6. "Become involved in their preferred ways"	6. Patient involvement in decision-making
Quality of the choice made:	
7. "Improve the match between the chosen option and the patients' values"	7.1. Knowledge 7.2. Accuracy in the patient's perceived probabilities 7.3. Concordance between chosen option and patient's values

The different elements of decision quality of the IPDAs definition have been measured in studies evaluating PDA effectiveness using different measurement instruments such as the Preparation for Decision Making Scale, the Decisional Conflict Scale and the Control Preferences Scale [61]. There is considerable consensus that PDAs should be evaluated according to the IPDAS criteria of PDA effectiveness [61].

There are other definitions of decision quality. For example, Sepucha et al. [68] have defined decision quality as "the extent to which a decision reflects the considered preferences of a well-informed patient, and is implemented" [55, 68]. This definition emphasises similar outcomes of quality of the choice made as does the IPDAS definition (namely, knowledge outcomes and concordance between the chosen option and the patient's values), but it makes an additional decision quality requirement: that the choice is implemented. The team led by Dr Sepucha in the Health Sciences Centre of the Massachusetts General Hospital has developed a set of Decision Quality Instruments (DQI) for a number of health care conditions such as back pain or breast cancer [69]. Each DQI, completed by a patient, assesses 1) the knowledge of the patient, 2) the concordance of the patient's choice with the patient's values, and 3) the extent to which the patient was involved in shared decision-making with his/her clinician [69].

Kaltoft et al. [21] have criticised existing measurements of decision quality (including the DQIs) mainly on the grounds that they do not incorporate patients' preferences for the different aspects of decision-making [21]. They have proposed to measure decision quality using a tool called My Decision Quality (MDQ) [21]. MDQ asks the patient, after making a health care decision, to first rate and then assign preference weights to the following eight decision quality criteria: 1) clarity about the options, 2) clarity about the likely effects and consequences of the options, 3) clarity about the importance of the effects and consequences of the options, 4) clarity about the chances of these effects and consequences, 5) trust in having received the best possible information, 6) receiving the desired level of decision support, 7) feeling in control of the decision to the desired extent, and 8) committing to implementing the choice [21]. MDQ, implemented using the multi-criteria decision analysis (MCDA) software *Annalisa in Elicia*, combines a patient's ratings and preference weights over the above eight criteria into an overall score of decision

quality which reflects how important each aspect of decision-making is for him/her personally [21].

4.1.4. Health outcomes.

Health outcomes include general health status (e.g. as measured by the 36 item Short Form Health Survey or by the EuroQoL EQ-5D instrument), condition-specific health outcomes (such as back pain, angina or bleeding), depression, anxiety and regret [5]. Some authors have argued that considering the main goal of health care is to improve health, the main measure of effect of PDAs should be their impact on health outcomes [70, 71] . Other authors argue that judging the success of PDAs primarily by whether or not their use leads to better health outcomes is inappropriate [61, 64]. This is so because: 1) PDAs are used in preference-sensitive decision situations involving the choice between alternative health care options none of which is clearly superior to the others in terms of health outcomes, 2) these decisions are made under conditions of uncertainty, which implies that beneficial or adverse health outcomes may be due to chance and 3) focusing on improving health outcomes ignores the possibility that patients may value other outcomes more than health outcomes [61].

4.1.5. Outcomes related with patient behaviour after using the PDA.

Behaviour-related outcomes include 1) the option implemented by the patient and 2) patient adherence to the chosen option [5]. The main criticism made with respect to using this type of outcomes as a basis for measuring the effectiveness of PDAs is that PDAs are not intended to promote one health care option over another; instead, the appropriate choice of option for a patient will depend on his/her preferences for the different consequences of these options [52].

4.1.6. Health systems outcomes.

These outcomes include the length of the consultation using a PDA, the costs incurred as a results of using a PDA, and the cost-effectiveness of using a PDA [5]. Including these outcomes as measures (but not as sole measures) of PDA effectiveness is important from the perspective of the health service provider [61]. This is so because if health systems are to fund access to PDAs, it is important to know the impact of these tools on costs and cost-effectiveness [61].

4.1.7. Conclusion.

As has been shown in the previous sections, there is debate in the literature about what is the most appropriate measure of effectiveness of PDAs. The definition of the IPDAS PDA effectiveness criteria, which results from a consensual effort by researchers, PDA developers and other stakeholders to develop common standards for the evaluation of PDAs, indicates that there is considerable agreement that, notwithstanding other goals (such as improving health), PDAs primarily aim to support decision processes which are conducive to patients making decisions that are consistent with their informed preferences.

5. What should a good PDA contain and how should it be developed?

5.1. IPDAS criteria to assess the quality of the PDA content and of the PDA development process.

Along with the IPDAS criteria of PDA effectiveness, the IPDAS Collaboration developed a consensus checklist of criteria for assessing the quality of the content of PDAs and the quality of the PDA development process [67].

Quality criteria regarding the content of PDAs are grouped into four dimensions:

1. “Provide information in sufficient detail for decision making” [67]. Patients should be provided with sufficient information to generate the knowledge that allows them to make preference-based choices [72]. Such information should include a description of the condition and the health care options involved, a description of how the condition would evolve if it is not treated, the likely benefits, the likely harms, and probabilistic information about both [72]. Other patient information needs should be identified [72];

2. “Present probabilities of outcomes in an unbiased and understandable way” [67]. Trevena et al [73] developed an expert consensus summary of relevant issues in the presentation of probabilistic information of the options’ consequences in PDAs, as well as evidence-based guidance on how to best address these issues. The issues and guidance presented in the summary are numerous and are described in section 6.5.1 of this review.

3. “Include methods for clarifying and expressing patients’ values.” [67] Value clarification methods aim to help patients to, *inter alia*: retrieve their preferences for relevant option consequences that may be overlooked, make holistic comparisons between options, and make preference trade-offs between option consequences [74]. Value clarification methods can be implicit or explicit [75]. Implicit value clarification methods are not interactive and include 1) vivid descriptions of the physical or mental impact of the consequences of the options and 2) other patients’ testimonials of their attitudes towards the options’ consequences [75]. Explicit value clarification methods are interactive and include 1) utility-based techniques [75] such as the standard gamble [76] and 2) non-utility-based approaches [75] (e.g. comparing the pros and cons of the options [77], the balance technique [78], the time trade-off technique [79], conjoint analysis [80], rank ordering the importance of consequences [77], or assigning importance weights on a visual analogue scale [77]).

4. “Include structured guidance in deliberation and communication” [67]. This dimension includes coaching (support through decision-making by a trained individual) and guidance (an explicit component of the decision support material aimed at facilitating the decision process) [81]. Coaching and guidance can aid patients to reflect on the steps of making a decision, anticipate and help to prevent cognitive biases in patients’ judgments, and engage patients in the process of learning about the decision [81]. Coaching and guidance can also increase patient-clinician communication, e.g. by helping patients prepare questions or by training patients to raise difficult issues [81].

Quality criteria regarding the PDA development process are grouped around six dimensions:

1. “Presentation of information and options in a balanced manner” [67]. The presentation of information in PDAs is balanced if it is complete (including all relevant options and all relevant option-related information such as possible benefits and harms and probabilistic information), if it is non-directive (i.e. giving equal weight to positive and negative aspects), and if it avoids bias in the processing of information by the patient [82]. Balanced presentation of information in PDAs is important, among other reasons, to avoid cognitive biases that may affect patient

knowledge and perceptions of benefits and harms, and ultimately their preferences [82];

2. “Systematic PDA development process” [67]. To inform this dimension, Coulter et al [83] suggested, based on a literature review of previous experiences that a systematic development process for a PDA should include the following steps: 1) a scoping stage to determine the decision, the target audience and the theoretical framework informing the PDA; 2) a design stage to determine patient needs for information, the format and content of the PDA and how the PDA will be delivered, 3) an alpha-testing stage to determine the comprehensibility and usability of the PDA, 4) a beta-testing stage to determine its feasibility in real practice [83];

3. “Use scientific evidence that is cited in a reference section” [67]. Informed consent requirements, quality of care principles and evidence-based medicine all suggest that the development of a high quality PDAs should incorporate comprehensive, up-to-date and critically appraised (“best available”) evidence on the options and on the probabilistic information regarding the consequences of these options for patients [84];

4. “Disclose conflicts of interest” [67]. PDAs should report prominently and in clear language 1) sources of funding for their development and distribution and 2) whether the people or institutions involved in the development of the PDA can incur in gains or losses derived from the choices patients make using the PDAs [85]. In addition, PDAs should report that no organisations with a commercial interest in the options included in the PDA have provided funding for the development and distribution of PDAs [85];

5. Quality criteria for PDAs including patient stories [67]. Patient stories (e.g. testimonials of other patients, or narratives of health practitioners discussing the experience of patients making the same decision [86]) can be a useful means to communicate decision-related information in PDAs [86]. The content of patient stories included in PDAs should report both positive and negative experiences [67]. PDAs should not include testimonials (or other narratives) of patients without their explicit consent and should report the reasons (financial or not) for patients agreeing to share their stories [67];

6. Quality criteria for PDAs delivered online [67]. PDAs delivered over the internet should 1) be navigable one step at a time, 2) be easy to return to if navigated away from, 2) permit searches for key words, 3) provide safeguards for (and feedback on) any personal health data entered into them, and 4) be printable as a sole document [67].

The IPDAS checklist is being promoted as guidance for the development of PDAs [87] and is increasingly used by PDA developers to inform the development and evaluation of PDAs [87]. An instrument (IPDASi [88]) is available to assess quantitatively the different IPDAS quality dimensions.

5.2. Criticisms of the IPDAS quality checklist.

Some criticisms have been made of the IPDAS checklist of PDA quality criteria. McDonald et al [89], focusing on the quality dimension "presenting probabilities in an unbiased and understandable way", found that 1) some of the key concepts underlying this dimension have not been defined and that 2) that there are gaps in the empirical evidence and in the theoretical basis supporting the use of specific criteria within this dimension [89]. Bekker [87] has pointed out that the IPDAS checklist was developed by expert consensus as "much of the evidence-base to underpin each IPDAS domain was weak" [87]. In particular, Bekker argues that there is insufficient evidence about what are the active components of PDAs which facilitate decision-making [87]. In this sense, different theories of decision-making emphasise different active components [87]. Elwyn et al. [90] illustrate this debate by considering how different theories of decision-making can inform the design of PDAs [91]. For example, the design of a deliberation component (the element of PDA that supports patients in deliberating about their choice [90]) would differ across theories of decision-making (e.g. Prospect Theory [91], Fuzzy-trace Theory [92] or the Rational-emotional Theory of Decision Avoidance[93]) [90].

5.3. Conclusion.

The development by the IPDAS Collaboration of a consensus checklist to assess the quality of PDAs was prompted by concerns regarding the quality of existing PDAs [30]. The checklist provides agreed standards for assessing the quality of the content and of the development process of PDAs, and is increasingly used by PDA developers. In 2014, the IPDAS checklist was proposed as the basis for a set of

standards for the certification of PDAs [94]. However, it should be used with caution [87]. First, the evidence underlying the recommendations of the checklist is not entirely clear, as shown by McDonald et al [89]. Second, the checklist does not adequately reflect the existing debate in the decision sciences regarding what are the active components of PDAs which facilitate decision-making, as discussed by Bekker [87].

6. Evidence of the effectiveness of PDAs.

6.1. The Cochrane systematic review of trials evaluating PDA effectiveness.

The Cochrane Collaboration published in 2014 the latest update of a systematic review of clinical trials evaluating the effectiveness of PDAs [5]. This review synthesised the evidence of PDA effectiveness for 115 studies covering 46 health conditions [5].

From the review, PDAs have proven to have a number of benefits on variables mapping onto the IPDAS criteria of PDA effectiveness [5]. First, compared to standard consultations the use of PDAs leads to patients having higher levels of knowledge with respect to the options and their outcomes [5]. In addition, the use of PDAs which include the probabilities of the consequences of the options leads to an increase in the accuracy of the perceived probabilities by patients [5]. These two results highlight the limitations of standard clinical consultations in terms of providing the necessary information for helping patients make preference-sensitive decisions [5]. Third, the use of PDAs *versus* standard consultations results in lower decisional conflict with respect to feeling uninformed and feeling unclear about values [5]. Fourth, compared to standard consultations, the use of PDAs leads to a reduction in practitioner-controlled decision-making: patients who use PDAs have a higher level of involvement in making decisions [5]. Fifth, compared to standard consultations, using PDAs leads to higher levels of consistency between the options chosen by patients and patients' values [5]. The review states that, in addition, PDAs with explicit value clarification exercises lead to an increase in values-based choice compared to PDAs without these exercises [5]: however, this is not entirely clear, as some of the studies reporting this effect use a variation of the Multidimensional Measure of Informed Choice [95] which does not measure directly the consistency between patients' values and their choices [61]. Other variables of proven PDA

effectiveness include overall decisional conflict, proportion of patients undecided and patient-practitioner communication [5].

PDAs have limited effects in terms of increased satisfaction, either with the decision or with the decision-making process [5]. Also, PDAs have limited effects in terms of improved health outcomes, which is unsurprising given that they are often used in preference-sensitive decision situations where no option has a distinct advantage over other options [5]. While the effects of PDAs on the options chosen by patients and on the length of the clinical encounter are variable [5], little is known about how the use of PDAs impacts on other measures of PDA effectiveness such as adherence to the chosen option or cost-effectiveness [5].

Although the 2014 update of the Cochrane systematic review provides an in-depth analysis of the effectiveness of PDAs, it does not distinguish between different PDA formats, different modes of PDA delivery, or different theoretical frameworks underlying the development of PDAs. With respect to specific PDA components, apart from describing the benefits of 1) including probabilities of the consequences into PDAs and of 2) including explicit value clarification exercises into PDAs, the Cochrane review only provides additional results of PDA effectiveness for studies comparing detailed with simple PDAs. From the review, compared to simple decision aids, the use of detailed decision aids results in 1) higher levels of knowledge of the options and outcomes and 2) lower levels of overall decisional conflict and of decisional conflict related to feeling uninformed [5].

6.2. Effectiveness of different PDA formats.

In terms of the effectiveness of PDAs by different formats, two reviews of the literature were found: a review by Hoffman et al. [96] exploring the effectiveness of PDAs delivered over the internet and a systematic review of the effectiveness of computerised PDAs (which include internet applications) by Sheehan et al. [51].

The review by Hoffman et al. [96] found some evidence that internet-delivered PDAs led to an increase in knowledge, improved preparation for decision-making and lower decisional conflict, as well as, for screening PDAs, an increase in the likelihood of screening [96]. However, these results were based on few clinical trials which did not control for potential confounders (e.g. whether or not patients had

previous experience of decision support) and hence should be interpreted with caution [96].

The systematic review by Sheehan et al. [51] provides some insight on the effectiveness of computerised PDAs (including internet applications) compared with simpler PDAs. The results of this review suggest that the use of computerised PDAs 1) lead to a very small effect on increasing knowledge of the options and outcomes unless they include feedback and self-test features and 2) lead to a lower level of decisional conflict with respect to feeling uninformed and unclear about values immediately after using the PDA (although not after 3 months) [51].

6.3. Effectiveness of different modes of PDA delivery.

No reviews were found exploring the effectiveness of PDAs by mode of delivery. One clinical trial has explored the relative effectiveness of giving patients 1) a PDA and 2) an information pamphlet under two distinct delivery modes: A) when both the PDA and the pamphlet are delivered by a clinician during the consultation and B) when a researcher delivers both the PDA and the pamphlet before the clinical encounter [97]. The study found that, compared with patients getting the pamphlet, those patients to whom the PDA was delivered during the clinical consultation had a higher level of knowledge than those patients to whom the PDA was delivered before the clinical encounter [97]. The authors point out among the limitations of the study that the PDA was designed for delivery during the consultation, so the results may not apply to PDAs designed for use outside the consultation [97].

6.4. Effectiveness of PDAs with different underlying theoretical frameworks.

Durand et al [26] identified and described the theoretical frameworks of PDAs included in the 2009 update of the Cochrane systematic review of PDA effectiveness. Although they did not undertake a comparative analysis of PDA effectiveness by type of framework, they found that two thirds of PDAs included in the Cochrane review were not based on any theory or model of decision-making [26]. Of those PDAs that were, there was little evidence that the design, development and evaluation of these PDAs was in accordance with the theory or model on which they were based [26]. Only one review was found of studies explicitly comparing the relative effectiveness of PDAs by type of underlying theory or model of decision making. This was a systematic review (previously mentioned in this text) of the

effectiveness of computerised PDAs [51]. In this review, the authors did not find any unequivocal evidence that the use of theoretically-based computerised PDAs, compared to a-theoretical PDAs can lead to increased levels of effectiveness [51].

6.5. Effectiveness of specific components of PDAs.

6.5.1. Effective presentation of probabilistic information.

Section 6.1. showed that there is evidence that the inclusion of probabilities in PDAs leads to an increase in the accuracy of the perceived probabilities by patients [5]. Trevena et al [73] developed an expert consensus document distilling the evidence regarding best practices in the presentation of probabilistic information in PDAs. A summary of this evidence includes the following points:

1. Appropriate numeric formats for the presentation of probabilistic information differ across tasks [73]. To illustrate this point with four examples:

A) for the task of presenting probabilistic information for two events which are independent (e.g. the probability of migraine symptom relief within 24 hours with treatment versus placebo) percentages (e.g. 90% versus 0%) are easier to understand than simple frequencies (e.g. 90 in 100 versus 0 in 100) [73];

B) for the task of presenting changes in probabilistic information (e.g. changes in the probability of disease before and after treatment), the absolute risk reduction (the probability of disease after the treatment minus the probability of disease before the treatment) is preferable to the relative risk reduction (the absolute risk reduction divided by the probability of disease before the treatment), as this last measure tends to bias (magnify) patients' perceptions of the change in probabilities and is not easy to understand [73];

C) for the task of presenting probabilistic information about connected events (e.g. the probability of disease given an irregular test result), natural frequencies (a step-by step description of the calculation of the posterior probability of an event based on the way individuals would learn it in real life [98, 99]) are proposed as the best format to help patients understand these probabilities [73];

D) for the task of presenting probabilistic information regarding options' consequences that occur over time, different approaches can be used when data is

available (which is usually a problem [73]), such as presenting such probabilistic information at one point in the future (e.g. the probability of cardiovascular disease in 10 years if a particular treatment is taken), at several points in the future, presenting the cumulative probability of the consequences over a patient's lifetime, or presenting survival and/or mortality graphs [73].

2. The reference class (i.e. the denominator of the probabilistic information presented) should be defined and used consistently throughout the PDA [73];

3. Presenting probabilistic information about the context (e.g. the probability of death from major causes other than the condition at hand) helps patients get a wider perspective of the risk of disease [73]. Labels qualifying probabilities (e.g. how "bad" a particular probability is) should be used with care, as they can affect patients' perceptions of risk [73];

4. There is little consensus about the best way to communicate uncertainty around probabilistic estimates [73]. Representations of randomness (the unpredictability of future outcomes) [100] can be made using icon arrays showing the number of individuals affected in a scattered rather than sequential fashion [73, 100]. These representations do not seem to significantly affect patients' perception of risk [73], but it is not clear whether or not they increase patients' understanding of uncertainty [73]. Representations of ambiguity (uncertainty about how reliable, credible, or adequate the probabilistic information is [73, 101]) can be made presenting confidence intervals around probability point estimates [73], e.g. via textual statements or visual aids (solid or blurred bars) [101]. The communication of ambiguity in PDAs may lead to ambiguity aversion [73];

5. Visual displays of probabilistic information may help in increasing the accuracy of the probabilistic information perceived by patients [73]. The types of visual displays which are more easily understood by patients include vertical and horizontal bars as well as pictographs (i.e. icon arrays) [73];

6. The impact of tailoring probabilistic information to individual patient characteristics (e.g. by whether or not that particular patient has specific risk factors) on PDA effectiveness is not clear [73];

7. The effect of interactive features for presenting probabilistic information in web-based PDAs is not well known [73].

6.5.2. Effectiveness of PDAs using value clarification methods.

Section 6.1 reported the evidence from the Cochrane systematic review of trials of PDA effectiveness regarding the positive impact of explicit value clarification exercises on value-choice consistency [5] and the problems with this evidence [61]. Fagerlin et al. [74] reviewed, based on the results of an previous, unpublished systematic review [102], the empirical evidence of the effects of including value clarification methods in PDAs. This study showed that:

1. The most frequently used value clarification methods were comparisons of the pros and cons of options (46% of studies), utility elicitation (18% of studies), prioritisation (i.e. rank ordering of consequences) (11%) and rating scales (11%) [74];

2. Only 13 trials compared the effectiveness of PDAs with value clarification exercises with PDAs without value clarification exercises [77]. Decision processes improved in five trials, but other PDA effectiveness outcomes were not measured with enough frequency to draw conclusions about the positive impact of value clarification methods on PDA effectiveness [77].

6.5.3. Effectiveness of PDAs using personal stories.

Bekker et al. [86] reviewed the evidence regarding the comparative effectiveness of PDAs with and without patient stories. The review found that the inclusion of patient stories in PDAs has an effect on both 1) deliberative and 2) heuristic (i.e. intuitively-experiential) strategies of information processing in patients [86]. With respect to the deliberative strategies, the review found, *inter alia*, an increase in patients' perceptions of making 1) informed decisions and 2) decisions based on their own values, as well as more stable choices, knowledge and preferences over time [86]. With respect to the heuristic strategies, the review found, *inter alia*, a decrease in counterfactual reasoning and an increase in emotional perceptions (e.g. fear of illness) [86]. However, the review found insufficient evidence about what are the active ingredients in personal stories which may facilitate decision-making [86].

6.5.4. Effectiveness of including structured guidance and coaching in deliberation and communication in PDAs.

Stacey et al. [81] reviewed the evidence available regarding the effectiveness of including coaching and guidance in PDAs. No trials were found isolating the effectiveness of guidance in PDAs [81]. Coaching added to a PDA improved knowledge and decreased costs compared to standard consultations [81]. Only four trials compared 1) coaching plus a PDA with 2) a PDA alone [81]. Hence, the evidence of the impact of coaching on the effectiveness of PDAs is limited [81]. Within these trials, one study found that the addition of coaching to a PDA increased participation in decision-making and decreased anxiety [103] and another study found that adding coaching to a PDA decreased costs [104].

6.6. Conclusion.

There is substantial evidence that PDAs, compared to usual care (i.e. standard consultations), are effective in terms of improving a number of criteria of quality of the decision-making process (e.g. decisional conflict, patient involvement in decision-making) and in terms of improving key elements of the quality of the choice made (knowledge outcomes and value-choice consistency) [5]. There is evidence that, compared to simpler PDAs, more detailed PDAs have a positive effect on knowledge and decisional conflict [5]. There is also evidence that the inclusion of probabilities in PDAs increases patient knowledge [5]. The literature on risk communication provides a number of evidence-based recommendations about best practices in the presentation of probabilistic information in PDAs to improve patient understanding of this information [73].

It is unclear whether or not the inclusion of value clarification methods in PDAs leads to improved decision-making process [74] or to increased value-choice consistency [61] compared to PDAs that do not include these exercises. It is also unclear whether including patient stories in PDAs facilitates decision-making [86], and whether or not guidance and coaching increase PDA effectiveness [81]. Finally, it is not clear whether the use of theoretically-based PDAs, such as MCDA-based PDAs, leads to increased levels of PDA effectiveness compared to a-theoretical PDAs.

7. Implementing PDAs in practice.

Although PDAs have shown to be effective across a number of outcomes [5], their implementation in routine clinical practice has been achieved to a level that is "less than expected" [105]. PDA implementation is challenging [7, 105]: indeed, a number of implementation barriers have been identified in the literature that hamper the widespread adoption of these tools. These barriers, and suggested solutions, include:

1. Clinicians' concerns about the content of PDAs [7, 105], including concerns that 1) this content is not comprehensive enough and that 2) it is out of date [7, 106]. Ensuring the comprehensiveness of the content of PDAs can be achieved by promoting the uptake of PDA quality standards [7], such as those included in IPDAS Collaboration checklist [7, 66]. With respect to the outdatedness of PDAs, it is clear that new evidence regarding the consequences of the options is quickly available for many clinical decisions and it needs to be regularly incorporated into existing PDAs [7]: stronger connections between PDA developers and those who generate, synthesise and analyse the evidence can help achieve this aim [7];

2. Lack of awareness in physicians regarding the availability of PDAs [7] and lack of training in the use of PDAs and Shared Decision Making (SDM) [7, 105, 107]. Training in decision support may help overcome these barriers [7, 108]. An environmental scan of training programmes in SDM [109] found, between 1996 and 2011, that 54 SDM programmes (most often targeting licensed health care practitioners) had been put into practice [109]. These programmes varied substantially in their content and were often not evaluated [109], suggesting the need for accreditation mechanisms [107];

3. Competing demands and time constraints on physicians [105]. These constraints limit the success of PDA implementation strategies that depend on clinicians identifying patients for using these tools [105]. One more successful strategy is to use system-based approaches for distributing PDAs to patients [105], such as 1) mailing PDAs to all eligible patients or 2) getting practitioners to hand PDAs to eligible patients to be filled at home [110]. Studies using this strategy have still found large differences between patients identified as eligible and patients actually using PDAs [105].

4. Clinicians' concerns about the adequacy of PDAs for some groups of patients [7]. Clinicians perceive that different groups of patients have different abilities and/or willingness to use PDAs [7] and to engage in SDM [107, 111]. In fact, vulnerable patients (such as elderly patients or patients with little education), compared with other patients, report less interest in participating in SDM [107]. PDAs aiming to provide decision support to these patients should address their health literacy needs [112]. Successful strategies used in the design of PDAs to improve understanding in patients with low levels of health literacy include 1) presenting numerical data in tabular or graphical formats instead of text and 2) including videos to reinforce the message presented with verbal narratives [112].

Important facilitators for the implementation of PDAs identified in the literature include 1) training practitioners in the use of PDAs, 2) the (already mentioned here) availability of system-wide methods to initiate PDA use in patients and 2) the availability of a "clinical champion" to lead on implementing these tools [105].

8. Multi-Criteria Decision Analysis (MCDA) and medical decision making.

Multi-Criteria Decision Analysis (MCDA) [113], Multiple Criteria Decision Analysis [9, 114], Multi-Criteria Analysis [10], and Multiple-Criteria Decision Making [115] are all terms that make reference to a collection of formal methods that can be used to solve decision problems involving multiple, often conflicting, criteria. In this thesis the term Multi-Criteria Decision Analysis (MCDA) is used to refer to such collection of methods. In MCDA, an initial distinction can be made between 1) decisions involving the appraisal of a finite number of options and 2) decisions involving the design of optimal options from an infinite potential set [10]. This thesis focuses on MCDA methods that can be used to make decisions involving the choice of one alternative from a set of finite, well-defined options, as these are the kind of decision situations typically confronted by patients making health care decisions.

8.1. Overview of MCDA methods.

The role of MCDA methods in supporting clinical decision making has been advocated by several authors, e.g. Dolan [116] and Dowie et al. [19]. MCDA methods are designed to provide decision support in complex circumstances like those arising in many typical patient management decision situations [116], i.e. decision situations where the choice of option depends on making trade-offs between

the consequences of the options [116]. MCDA has been identified by Durand et al. [26] as one of the several decision-making theories informing the design of PDAs.

8.1.1. Steps of the MCDA process.

MCDA methods provide guidance to the decision maker(s) in exploring a particular decision problem [10]. When applied in full, the MCDA process involves the following eight steps [10]:

1) Establishment of the decision context [10]. An important element of establishing the decision context involves determining who is responsible for making the decision [10]. Is it a single individual? Is it a group of individuals? If the second is the case, it will be appropriate to use applications of MCDA which are suited for group decision-making, e.g. those reviewed in Kilgour et al. [117]. Another important element of establishing the decision context is understanding the main overarching objective of the decision [10]. Establishing the main objective is important in order to be able to establish lower sub-objectives which will be the basis for defining the measureable criteria on which the performance of the options can be evaluated [10]. Value-focused thinking [118] is a good approach to facilitate this process.

2) Identification of the options [10]. Options are important insofar as they are of value to achieve the objectives of the decision [10] and so, when they are not given, they should be established after the objectives of the decision are developed [10, 118]. This point is of limited relevance for MCDA in patient decision support using PDAs, where the health care options (including doing nothing) are normally given.

3) Identification of the objectives and of the criteria which show the value of the options' consequences [10]. Establishing the criteria on which to measure the performance of the options is an important part of MCDA. For this task it is useful to first determine the fundamental (i.e. important in themselves) sub-objectives to achieve the main decision objective and structuring these in a hierarchy or "value tree" [10], as for example is done using the value-focused thinking approach [118]. The value tree of criteria measuring these objectives, should:

- Be complete, i.e. it should include all the relevant criteria for making the decision [10];

- Be of adequate size, i.e. include as few criteria as possible so that it is "no larger than it needs to be" [10];

- Be non-redundant (i.e. it should not include criteria which are not important in making the decision) and not incur in double-counting of criteria [10];

- Be operational, i.e. all the criteria included in the value tree should be defined clearly enough to be assessed from both an objective and a value perspective [10];

- Adequately factor-in the impact of time, e.g. by defining the criteria with respect to a specific time horizon [10];

- Be such that all the criteria are mutually preferentially independent, i.e. the preference of the decision maker over different levels of performance on each criterion should not depend on the performance levels of the other criteria [10].

4) Description of the expected performances of the options on the criteria and- in the case of those methods involving steps 5) and 6) below - calculating scores [10]. The variables measuring the performances of the options on the criteria can be, for example, continuous quantitative (e.g. cost of treatment in GBP), categorical (e.g. presence or absence of side effects), ordinal (e.g. low/medium/high efficacy), and interval (e.g. three levels of cost of treatment: "between 0 and 100 GBP"/ "between 100 and 500 GBP"/ "more than 500 GBP"). The performance levels of the options on each of the criteria are normally represented as cells in a performance matrix with options in rows and criteria in columns [10]. Scores normally reflect the value (e.g. on a scale between 0 and 100) to the decision maker of the different performance levels of each criterion. These scores can be calculated using different methods, e.g. estimating value functions or via direct rating [10]. The type of scales on which the scores are calculated (e.g. interval or ratio scales) differ depending on the MCDA method. The extremes (the two points representing respectively a value of "0" and a value of "100") of the score scales used can be assigned by reference to the local set of options (the options taken into account by the decision-maker) or to a global set of options (i.e. a set of options including options not taken into account by the decision maker) [10];

5) Assignation of weights (N.B. or other inter-criteria parameters) reflecting the relative importance of the criteria [10]. The calculation of criteria weights (and other inter-criteria parameters) differs across MCDA methods;

6) Integration of the performances and the weights to obtain an overall value for each option [10]. Different MCDA methods perform this integration in different ways;

7) Examination of results [10], e.g. assessing the overall ranking of the options and any other information that is relevant for making recommendations about the relative performance of the options on the criteria;

8) Sensitivity analysis [10] e.g. to assess how changes in imprecise criteria weights affect the final results.

8.1.2. Types of MCDA methods.

A possible typology of the many MCDA methods available includes the following:

1. MCDA methods based exclusively on the analysis of the performance matrix [10, 119]. These basic forms of MCDA are non-compensatory, i.e. for a specific option they do not allow for bad performance in one criterion to be compensated by good performance in another criterion. They include: 1) dominance analysis, which involves eliminating dominated options, i.e. those options which perform no better than any other options on any of the criteria and worse than all other options on some criteria [10]; 2) disjunctive/conjunctive selection procedures, which respectively eliminate options if they fail to reach a threshold level on some or all the criteria and include options for further consideration if they reach threshold levels on some or all the criteria [10]; 3) the lexicographic ordering approach, which involves selecting the highest performing option on the most important criterion unless there is more than one option with the same highest performance, in which case the process is repeated with this subset of options for the next most important criterion until either there is only one option left or there are no more criteria left [10]; 4) elimination by aspects [10], which sequentially eliminates all options but one based on whether or not the options meet specific aspects, i.e. values or characteristics related with the criteria (e.g. when buying a car, a threshold lower price, or automatic versus manual transmission [120]) or arbitrary features that do not correspond to any of the criteria

[120]. All of these methods very much limit the incorporation of the decision maker's preferences into the decision [10];

2. Full aggregation MCDA approaches [113]. Typically, the way these approaches operate involves first scoring each option on each criterion and then combining these scores into a global score for each option. These approaches are fully compensatory [113]. In this group are such well-known methods as Multi-Attribute Value/Utility Theory (MAUT/MAVT) [11] and extensions of MAVT based on linear additive multi-attribute value functions such as the Simple Multi-Attribute Rating Technique (SMART), SMART with swings (SMARTS), and SMART exploiting ranks (SMARTER) [121]. This group also includes other simple linear additive models such as the Simple Additive Weighting (SAW) approach; it also includes the Analytic Hierarchy Process (AHP)[122] and its extension the Analytic Network Process (ANP) [123], Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) [124], and the family of UTA ("UTilites Additives") methods [125];

3. Outranking MCDA methods [113]. These methods are based on making sequential pairwise comparisons of the options (taking into account their performances on the criteria) so as to determine what is the preference relation that can be established between every option pair [126]. In contrast with full aggregation MCDA methods, these approaches may lead to cases of incomparability between options, and hence require additional exploitation procedures to be able to rank the options [126]. These methods are partially compensatory [113]. The most widely used outranking methods are the Election et Choix Traduisant la Realité (ELECTRE) [13] family of methods and the Preference Ranking Organisation Method for Enriched Evaluation (PROMETHEE) [14] family of methods, but there are others [127];

4. Goal, aspiration or reference-level MCDA methods [113]. These approaches operate by first defining a goal, ideal or reference level on each of the criteria and then identifying the option which is closest to that level [113]. Such methods, under the general rubric of goal programming [128] are often used in design multi-criteria problems where the objective is to identify (i.e. design) an optimal option from an infinite (or very large) potential set of options. For choice multi-criteria decision problems where the options are pre-determined (the subject of this thesis), several

MCDA methods based on this logic are available. One of these, Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), is based on ranking the available options by calculating an index of closeness to an ideal solution [129]. Other methods of this type can be interactive, such as the Visual Reference Direction approach [130]. This approach proceeds in an iterative fashion [130]. At each iteration, the decision maker identifies one preferred option and his/her aspiration levels on the criteria; with this information, an achievement function (based on the reference direction, i.e. the vector connecting the preferred option with the decision maker's aspiration levels) is minimised, leading to the selection of a (smaller in each iteration) subset of the options available; the process is finished when the decision maker is unable to find a more preferred option [130]. Other aspiration-level MCDA methods which use achievement functions include approaches based on the objective ranking of options [131]. In these approaches, the aspiration levels of the criteria are not determined via the preferences of the decision maker, but as objectively as possible [131].

5. Other methods.

5.1. Fuzzy MCDA methods [132]. In set theory, "fuzziness" refers to classes of objects in which there are grades of membership to a set [133]. These grades of membership are expressed using membership functions, defined between 0 (when the object definitely does not belong to the set) and 1 (when the object definitely belongs to the set) [133]. "Fuzziness" can be used to explicitly model imprecision in information or in knowledge into MCDA. For example, membership functions can be used to account for imprecision in the measurement of criteria performance levels when these are defined qualitatively (e.g. "fair" or "good") [134]. Membership functions can also be used to incorporate imprecision in the assessment of preference weights [134]. Fuzzy MCDA methods differ in terms of the way they integrate performance levels and preference weights to obtain an overall rating for each option, but a common procedure to perform this integration is the weighted average sum method (i.e. multiplying the performance level of each option on each criterion by the normalised weight of that criterion and adding across criteria) [135]. Since the resulting ratings can be fuzzy, it is often necessary to use additional procedures to be able to rank the options from most to least preferred [135]. Fuzzy approaches have

been developed for a number of MCDA methods, e.g. the AHP [136], outranking methods [137] and MAUT [138];

5.2. Stochastic Multicriteria Acceptability Analysis (SMAA) [139]. SMAA is a family of MCDA methods which explicitly deal with uncertainty, imprecision and/or missing information in 1) the performance levels of the options on the criteria and in 2) the preference weights of the criteria. In SMAA, the decision problem is represented in a stochastic fashion, i.e. defining suitable joint probability distributions for the uncertain/imprecise/missing variables (namely, the variables measuring the performance levels of the options on the criteria and the criteria weights) [139]. A decision model is then assumed for ranking, sorting or classifying options [139]. For example, for ranking options, SMAA-2 assumes a multi-attribute value function [140]. Monte-Carlo simulations are then used to draw many samples from the joint probability distribution of performance levels and weights and to generate results in terms of the assumed decision model - results which can be summarised statistically and/or graphically to draw conclusions [139]. For example SMAA-2 calculates a rank acceptability index measuring the percentage of all combinations of preference weights which give a particular option a specific rank: the best options are those with high rank acceptability indexes for the best ranks [139]. Other descriptive measures provided by SMAA-2 are 1) the central weight vector (representing the preferences of a "typical" decision maker choosing a particular option) [139] and the confidence factor (the probability that an option ranks first when its central weight vector is selected) [139];

5.3. Dominance-based rough set approach (DRSA) [141]. DRSA approaches are based on modelling the preferences of the decision maker(s) in terms of "if…then…" rules [141]. To explain DRSA for a choice decision problem, it is useful to first explain DRSA for a classification problem, i.e. a decision problem involving the classification of options into different categories. For a classification problem, the DRSA analysis starts with a data table, a table with options in rows, criteria in columns and evaluations (either quantitative or qualitative) of the options on the criteria in cells [141]. The DRSA data table is different from a performance matrix in that it contains two types of criteria: 1) condition criteria (N.B. the individual criteria relevant for the decision) and 2) decision criteria (one for each decision maker involved in the decision) which provide comprehensive evaluations of each of the

options [141]. When the decision involves only one decision maker (the example discussed here), then there is only one decision criterion. A key element of DRSA is the dominance principle, which states that if one option is at least as good as another option on a subset of condition criteria, then the first option should have a comprehensive evaluation (on the decision criterion) which is at least just as good as that of the second option [141]. DRSA uses rough sets (i.e. approximations of crisp sets, where a crisp set is a conventional collections of objects) and the dominance principle in order to derive "if…then…" decision rules [141], or decision rules linking the evaluations of the options on the condition criteria with their comprehensive evaluations on the decision criterion. For an example adapted from Greco et al [141], one such rule might be: "if the evaluation of an option in criterion 1 is at least 'fair' and its evaluation on criterion 2 is at least 'fair', then the option is comprehensively evaluated as at least 'fair' (i.e. 'fair' or 'good')". To modify DRSA to help solve choice multi-criteria decision problems, the data table needs to include pair-wise comparisons of options' evaluations and the dominance principles needs to be defined with respect to these pair-wise comparisons [141];

5.4. Verbal decision analysis (VDA) [142]. In choice decision problems, VDA (as implemented by the most well-known VDA method, ZAPROS [143]) ranks options by constructing a decision rule which is based on comparing verbal formulations (e.g. "below average", "average", above average" [142]) of the performance levels of the options on the criteria [142]. The procedure involves several steps. In the first step, the decision maker is asked to make trade-offs between all pairs of hypothetical criteria vectors where each vector contains the best possible performance level on every criterion but one. This allows, under certain conditions[1] [142], to construct a joint ordinal scale (a scale which provides a rank order of its elements and no information about the relative position of these elements) for these vectors [143]. For example, in a decision problem with two criteria, the joint ordinal scale might be: first ranked, hypothetical criterion vector ("average", "above average"); second ranked, hypothetical criterion vector ("above average", "average"); third ranked, hypothetical criterion vector ("below average", "above average"); fourth ranked, hypothetical criterion vector ("above average", "below average"). The next step in VDA involves considering the criterion vector of each real option and substituting

[1] Transitivity and mutual preferential independence of the decision makers preferences

the performance level of this option on each criterion by its rank order in the joint ordinal scale [143]. For example, consider two options: option 1 with criterion vector (“above average”, “below average”); option 2 with criterion vector (“average”, “average”). Substituting each performance level by its rank order on the joint ordinal scale yields respectively vectors (2,4) and (1,2). From this operation, it is clear that option 2 is preferred to option 1.

8.1.3. Conclusion.

In this overview of MCDA methods, the general process of MCDA has been identified. It is a step-wise process which guides the decision-maker through the steps of making the decision. The general steps of the MCDA process include: identifying the decision context, identifying the options, identifying and adequately structuring the objectives and their measureable criteria, developing the matrix of performance levels of the options on the criteria, integrating the information from the performance matrix to make a choice, analysing the results and performing sensitivity analysis.

In addition, a typology of MCDA methods has been presented and each type broadly described. A discussion of the relative merits of all these methods is beyond the scope of this review. However, the next section explores a selection of these MCDA approaches. This selection is based on 1) a number of long-established [126] full aggregation MCDA methods and a number of well-known outranking methods, 2) existing applications of these MCDA methods in the area of diagnosis and treatment of diseases, based on a recent systematic review and bibliometric analysis [144]. The selected MCDA approaches are:

1) Full aggregation MCDA methods.

1.1. Value and utility function methods, including: a) Multiattribute Value/Utility Theory (MAVT/MAUT), b) the Simple Multiattribute Rating Technique with Swings (SMARTS), c) the Simple Multiattribute Rating Technique Exploiting Ranks (SMARTER), d) Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH);

1.2. Simple Attribute Weighting (SAW);

1.3. The Analytic Hierarchy Process (AHP) and its generalisation the Analytic Network Process (ANP);

2) Outranking methods, including two of the most well-known [126]: a) Election et Choix Traduisant la Realité (ELECTRE) and b) Preference Ranking Organisation Method for Enriched Evaluation (PROMETHEE);

3) The Technique for Order of Preference by Similarity to an Ideal Solution (TOPSIS).

8.2. Exploring a selection of MCDA methods.

8.2.1. Value and utility function methods.

8.2.1.1. Multi-attribute Value/Utility Theory (MAVT/MAUT) and its extensions SMARTS and SMARTER.

MAVT/MAUT is a MCDA approach developed by Keeney and Raiffa [11] which is based on a number of axioms about rational decision making (for a description of these axioms, see Keeney and Raiffa [11]). A full aggregation MCDA approach, for a specific decision problem MAVT/MAUT combines 1) the performance levels of the options on the criteria and 2) the decision maker's preferences over the criteria via the construction of a multi-attribute value function, MAVF (under certainty in the consequences of the options) or a multi-attribute utility function, MAUF (under uncertainty in the options' consequences). Given a choice decision problem under certainty, each option *i* is assigned an aggregate score S_i on a MAVF. Similarly, for a choice decision problem under uncertainty, each option is assigned an aggregate score S_i on a MAUF. In each case, consideration of the aggregate scores of all the options allows for a full ranking of all these options from most to least preferred.

Multi-Attribute Value Theory (MAVT).

For a given choice decision problem under certainty in the performance levels of the options on the criteria, the main task in MAVT involves constructing a MAVF assigning a unique score to each option with multi-attribute consequences [11]. The MAVF is usually represented as a function of criterion-specific value functions. Depending on the structure of the decision maker's preferences, MAFVs may have different forms, e.g. multiplicative on the criterion-specific value functions or

additive on these functions [11]. Constructing a MAVF involves a sequence of tasks which can be quite complex, but which is substantially facilitated for an additive MAVF. In order to be able to assert that the preferences of the decision maker can be expressed with an additive MAVF, a particular assumption about the preference of the decision maker needs to hold, i.e. mutual preference independence of the criteria (MPIC) [145]. In essence, MPIC requires that any subset of criteria **C** is preferentially independent of its complementary set **C'**, i.e. that in any subset of criteria **C** the preferences of the decision maker between any pair of options differing on the levels of criteria in **C** do not depend on the levels of the criteria in **C'** [11]. If this condition holds, the aggregate score S_i of any option i on the (linear additive) MAVF can be written as:

$$S_i = \sum_{k=1}^{K} w_k v_{i,k} \qquad i = 1,2, \dots l \tag{1.1}$$

In equation 2.1, $v_{i,k}$ are the value scores of the local options (i.e. the options included by the decision maker as part of the choice decision) A_i $(i = 1,2, \dots l)$ on K single-criterion value functions scaled between 0 and 1, and w_k are K criterion-specific scaling constants adding to 1. The resulting MAVF is an ordinal MAVF which does not provide information about the strength of preference of the decision maker between the options. Note that mutual preference independence is a concept unrelated to statistical correlation between criteria. Statistical correlation is a measure of the extent to which two quantities fluctuate together. Two criteria can be correlated and be preference independent.

In order to construct the additive ordinal MAVF, the steps are [11]: 1) verifying that MPIC holds, which requires checking that the rates at which the decision maker substitutes the levels of one criterion for the levels of another criterion do not depend on the level of the remaining criteria, 2) constructing the single-criterion value functions using techniques based on indifference judgments such as the mid-value splitting technique [11], and 3) calculating the scaling constants using trade-off methods [11, 146]. Calculating the value scores of the options on the criteria and entering this information into equation 1.1 solves for S_i.

The additive MAVF is measurable, i.e. provides information about the strength of the decision maker's preferences over the options, if difference independence (DI) holds [145]. In essence, DI holds if, once it is verified that MPIC holds, the preference difference between any pair of options differing on the levels of one criterion do not depend on the common levels of the remaining criteria [145, 147]. The additive measurable MAVF has the same formula shown in equation (1.1).

The resulting measureable MAVF is built on an interval scale of measurement [9]: comparisons between the overall value scores of the options are made using differences in overall value. The first step in constructing a linear measureable MAVF is checking the required preference independence conditions. The second one is building K single-criterion value functions. Here, simpler methods than techniques based on indifference judgments can be used to build the single-criterion value functions. These methods include indirect assessment of the value function, constructing a value scale and direct rating [9]. The criteria weights w_k can be calculated, notwithstanding their calculation using trade-off methods (as for the additive ordinal MAVF) [145] using simpler methods, e.g. swing weights [9] or direct rating [147] (also termed "importance weights" [9]). Other approaches for weight elicitation include point allocation or rank order methods [146].

When the criteria are structured as a hierarchy or value tree, the simultaneous assessment of all the bottom-level criteria (e.g. via swing weights) can be a cumbersome task. This task can be simplified by assessing the weights at each level of the hierarchy.

A number of software tools can be used for decision support using MAVT, including: HiView [15], V.I.S.A. [148], and Logical Decisions [149].

Multi-Attribute Utility Theory (MAUT).

For a given choice decision problem under uncertainty in the performance levels of the options on the criteria, the main task in MAUT involves constructing a multi-attribute utility function (MAUF) [11]. The MAUFs (as the MAVFs in MAVT) are usually represented as functions on criterion-specific utility functions. In order for the decision maker's preferences to be represented by an additive MAUF, the assumption of additive independence of the criteria (AI) must hold [11]. To illustrate AI for the case of a decision problem with two criteria, AI holds if the decision

maker is indifferent between A) all lotteries involving with equal probability 1) variable levels x and y of two criteria and 2) arbitrarily fixed levels x' and y' of these two criteria and B) all lotteries involving with equal probability 1) variable levels of x on the first criterion and an arbitrarily fixed level y' of the second criterion and 2) an arbitrarily fixed level of x' on the first criterion and variable levels y of the second criterion [11]. In other words, AI holds if there is absolutely no interaction of preferences among criteria [150]. If AI holds, then:

$$S_i = \sum_{k=1}^{K} w_k u_{i,k} \qquad i = 1,2, \dots l \tag{1.2}$$

In equation 2.2, $u_{i,k}$ are values on K single-criterion utility functions and w_k are K scaling constants.

The steps for constructing a MAUF can be simplified as follows [11]: 1) defining the region of values over which the utilities of the criteria will be assessed, 2) setting up relevant lotteries to verify AI, 3) if AI holds, constructing single-attribute utility functions (e.g. using the variable probability method or the variable certainty equivalent approach [151]), 4) calculating the scaling constants using trade-off methods.

MAUT is supported by the Generic Multiattribute Analysis (GMAA) software tool [152].

Simple Multiattribute Rating Technique with Swings (SMARTS) and SMART Exploiting Ranks (SMARTER).

Edwards developed the Simple Multiattribute Rating Technique (SMART) method in part out of concern regarding the difficulty and instability of the indifference judgments required to construct multiattribute preference structures in Keeney and Raiffa [11, 121]. The original procedure used to calculate criteria weights in SMART (ranking the criteria in order of importance, assigning the most important criterion a reference weight of 100 and assigning weights to the other criteria relative to this reference weight) had the limitation that it ignored the ranges in the values of the criteria to determine criteria weights so SMART is not recommended any more

[121]. Two improvements have been suggested: SMART with Swings (SMARTS) and SMART Exploiting Ranks (SMARTER) [121]. SMARTS and SMARTER both provide simple approaches to building MAVFs based on *a "*strategy of heroic approximation*"* [121] which involves identifying "the simplest possible judgments that have any hope of meeting the underlying requirements of multiattribute utility measurement, and try to determine whether they will lead to substantial suboptimal choices in the problem at hand" [121].

In SMARTS, 9 steps are proposed for solving a multi-criterion decision problem [121]:

1) Establish the purpose of the value elicitation and identify the decision maker(s);

2) Construct a value tree. If more than one decision maker is involved in making the decision, a group exercise can be useful to get agreement from all the decision makers on the final structure of the value tree (if possible be limited in size to twelve criteria and avoiding criteria duplicates and criteria overlaps) and on the labels of the criteria (which should be unambiguous) [121].

3) Identify the options. If the options are not known in advance, the value tree can be used to generate a set of real or hypothetical options - it is useful to anticipate the range of performance levels of these hypothetical options on the criteria so it is not too narrow [121];

4) Construct the performance matrix. If physical measures can be used to measure the performance levels of the options on the criteria, they should be used [121];

5) Eliminate dominated options. When, after eliminating dominated options, the range of the performance levels on a particular criterion is greatly reduced, consider eliminating that criterion [121];

6) Calculate single-criterion utilities for the elements of the performance matrix. If possible, single-criterion utility functions linear on the performance levels of the criteria should be used, as they are easy to calculate [121]. To check the linearity assumption for a particular single-criterion utility function, a first check involves testing the monotonicity of the function eliciting internal maxima/minima [121]; a second check involves testing the curvature of the function, e.g. by eliciting the changes in value induced by small changes in the performance levels at several

points in their range [121]. If the linearity assumption does not hold, single-criterion utility elicitation methods which do not rely on making indifference judgments between lotteries can be used [121]. A final test in this steps involves checking for conditional monotonicity (CM) to verify that an additive value model on the single-criterion utilities is adequate [121]. CM is not met in instances in which the direction of the decision maker's preferences for the levels on one criterion changes for different levels of another criterion [121]. If CM is not met, a non-additive value model should be used [121].

Steps 7 and 9 involve the calculation of swing weights [121].

7) Rank order of the criteria weights. This step is implemented as follows [121]. First, selecting the most important criterion by imagining a hypothetical option which has no value (i.e. a value of 0) on all the criteria except on a criterion of choice, where it has the highest value possible (i.e. a value of 100). The criterion chosen by the decision maker is the most important. This procedure is then repeated for the remaining criteria.

8) Elicit swing weights using direct estimates of their magnitude and calculate the multiattribute utility of each option. To elicit the swing weights, the following procedure can be used [121]. First, assign the most important criterion a weight of 100. Next, elicit, for the second most important criterion, the worth (in terms of a number between 0 and 100) to the decision maker of a swing in value between 0 and 100 in this criterion compared with a swing in value between 0 and 100 in the most important criterion. The resulting number is the weight of the second most important criterion. The procedure is then repeated for the remaining criteria. To obtain the final weights w_k of the K criteria, they are normalised to add up to 100. To calculate the multiattribute utility of each option, the weights and the performance levels of the option on the criterion are combined using the relevant (i.e. either additive or non-additive) value model identified in step 6.

9) Decide.

SMART Exploiting Ranks (SMARTER) follows the same steps as SMARTS except for the elicitation of weights, i.e. step 8) [121]. In SMARTER, the weights w_k of the K criteria are rank order centroid (ROC) weights [121]. These weights are calculated

by solving a system of equations on the hypersurface (or simplex) of weights defined by these weights adding up to 1 [121]. The formula for each weight w_k is:

$$w_k = \left(\frac{1}{K}\right)\sum_{j=k}^{K}\left(\frac{1}{j}\right) \tag{1.3}$$

SMART with Swings (SMARTS) is supported by software applications such as Logical Decisions; SMARTER is supported by e.g. Web-HIPRE and Logical Decisions.

8.2.1.2. Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH).

The MACBETH approach separately elicits single-criterion value functions and criteria weights and aggregates both using an additive value model. The aggregate scores S_i of the options are defined on a cardinal value scale [153]. However, it uses a different approach to MAVT/MAUT for eliciting the inputs of the decision model.

The MACBETH approach to calculating each single single-criterion value function is based on making pairwise comparisons of difference in attractiveness between either 1) performance levels (expressed either qualitatively or quantitatively) on the criterion or, if the decision maker so wishes, 2) between the options themselves [113, 153]. For the remainder of this section, the assumption is that the performance levels are expressed qualitatively. The pairwise comparisons between the levels are done on a 7-point semantic scale of differences in attractiveness with the following categories [113]: 1) “none”, 2) “very weak”, 3) “weak”, 4) “moderate”, 5) “strong”, 6) “very strong”, 7) “extreme”. To construct each single criterion value function, it is useful to provide a description (in terms of their performance levels on that criterion) of two reference options [153], e.g. one “neutral” (neither satisfying nor satisfying) [153] and one “good” (undoubtedly satisfying) [153], which will serve to anchor the value scale and later to facilitate the calculation of the criteria weights [153]. Assuming this is done, the next step is to perform the pairwise comparisons of difference in attractiveness between pairs of performance levels. These pairwise comparisons can be performed in any order, although it is recommended to rank the performance levels in order of decreasing attractiveness [113]. Linear programming techniques are

used to calculate the single-criterion value function [153], which can have negative values and values higher than 100 if the reference options are not the worst and the best on that particular criterion [153].

The criteria weights are calculated based on pairwise comparisons of differences in attractiveness between the criteria [113, 153]. The performance levels of the reference options on the criteria (described above) can be used as the basis for eliciting swings in differences of attractiveness between criteria [153]. Linear programming is used to calculate the actual weights [113].

MACBETH can be implemented using the dedicated software M-MACBETH [154]. In MACBETH, it is important to verify that the judgments of the decision maker are consistent [113]. The software M-MACBETH highlights inconsistencies in judgements and supports the decision maker in correcting these inconsistencies.

As with other value function models presented here, MACBETH allows for structuring the decision problem in terms of a value tree, but unlike these, it does not allow for more than one parent criterion (i.e. a criterion composed of several sub-criteria) at the same level of the hierarchy [113].

8.2.1.3. Assessment of MAUT/MAVT and MACBETH.

Value function methods calculate the aggregate score S_i of each option A_i ($i = 1,2, \dots l$) as an overall score between 0 and 1 on a multi-attribute value (or utility) function (MAVF/MAUF). Such a function is built on an interval scale. Differences in overall option scores reflect differences in the strength of preference between options. To compare options, ratios of the overall scores of the options (corresponding to statements of the type "option A_1 is twice as preferred as option A_2 ") are not meaningful [9].

Riabacke [146], citing French and Rios Insua [155], described prescriptive decision analysis as a method which should be: 1) grounded on an axiomatic basis accepted by users, 2) feasible, i.e. practical in its implementation , 3) robust, in the sense that the sensitivity to changes in the inputs should be understood, 4) transparent to users, and 5) compatible with a wider philosophy, i.e. with the user's view of the context. The work by Keeney and Raiffa [11] resulting in MAVT/MAUT is an early prescriptive approach in the field of decision analysis with multiple criteria [146]. An

important strength of this approach is that it is based on constructing the MAVF/MAUF in a feasible way while respecting an axiomatic basis of rational decision making. However, as mentioned above, the elicitation procedures in MAVT/MAUT are still relatively complex and sometimes unstable. SMARTS and SMARTER use a *strategy of heroic approximation* to facilitate the MAVT/MAUT approach as much as possible while respecting its axiomatic basis. MACBETH was also developed in response to the complexities of the elicitation procedures required in MAVT/MAUT [124]. In MACBETH, the calculation of MAVFs, based on integrating a series of qualitative judgments of difference in preference (systematically checked for consistency) using linear programming techniques, eschews the complexity of the MAVT elicitation procedures proposed by Keeney and Raiffa.

In terms of the single-attribute value or utility functions, the elicitation of preferences is a task that is cognitively demanding and prone to error [146]. Different procedures for eliciting values/utilities (like the ones that have been described here) with the same normative basis should yield the same ordering of preferences (the assumption of procedural invariance [156]), but empirical studies have shown that this may not be the case [146]. Approaches have been proposed to choose the most appropriate preference elicitation method depending on the specific situation [157].

In linear additive MAVFs/MAUFs (a commonly used MAVT/MAUT model), the weight of a particular criterion is "a scaling factor which relates a unit on its measurement scale to a unit on the measurement scale for any other criterion" [158]. The elicitation of the criteria weights can be undertaken using 1) ratio weight approaches such as direct rating, swing weights, trade-off methods and point allocation [146] and 2) imprecise weight elicitation methods, e.g. rank order methods or methods based on semantic scales (as in MACBETH) [146]. When the decision problem is structured as a value tree, the weight of each criterion at an intermediate level of the tree is interpreted as the total weight of its sub-criteria [158].

Ratio weight methods are hard to obtain accurately [146]. In the elicitation of these types of weights, the range of the value scale on each criterion needs to be taken into account. When the top and bottom values of the value scales are fixed based on best and worst performance levels of the options available locally, the value scale is

termed a local scale [159]. When the top and bottom values of the value scales are fixed based on the best and worst values according to the decision maker's "experience, aspirations or imagination" [159], the value scale is termed a global scale [159]. Importance weights have been shown to be inappropriate with local scales in which the ranges of the values on the criteria are small and hence swing weights are advocated when these scales are used [159]. However, empirical studies have shown that people often do not adjust properly for the ranges in the criteria [146]. Several methods (including the generalised use of global scales [159]) have been suggested to improve this adjustment [146].

Within imprecise weight elicitation methods, two will be briefly discussed here with reference to the value function methods discussed above: rank order methods and methods based on semantic scales. Rank order methods, used e.g. in SMARTER, have been proposed on the basis that they adequately identify the best option between 75% and 85% of the time [10, 121]. The use of semantic scales with verbal terms in elicitation has the general problem that the same verbal expression may have different meanings for different people [146]: the interpretation of the numerical weights resulting from these verbal expressions can be difficult.

In summary, value function methods have the advantage that they are solidly grounded in axioms of rational decision making. Classic MAVT/MAUT methods are can be complex to use, even if the criteria are value/utility independent, but good prescriptive approximations are available (SMARTS and SMARTER). MACBETH is an alternative to MAVT that relies 1) on pairwise comparisons of difference in preference between options (or option performance levels) and criteria and 2) on an iterative, software-supported, consistency-checking process. The elicitations of values, utilities, and weights can be subject to error, so they should be done with care. Careful examination of results and their validity is important, as is conducting sensitivity analysis on the inputs of the decision model [10].

8.2.1.4. Application of value function methods in the area of diagnosis and treatment of diseases.

In their systematic review of MCDA applications in health care, Adunlin et al. [144] report several applications of value function methods in the area of diagnosis and treatment of diseases. Van Til et al. [160] studied the applicability of different weight elicitation techniques, including those of SMART and SMARTS, in patients with mild cognitive impairment being treated for a stroke-related disability. They found that these individuals are willing (as well as able) to use these weighting approaches in a decision situation, although they did not express a preference for a particular method [160]. Chang et al. [161] used multi-attribute utility model to assess the factors that predict womens' decisions to receive epidural analgesia. Pinheiro et a. [162] use MACBETH and Bayesian networks to determine what clinical aspects are most relevant in the diagnosis of Alzheimer's disease. Shaw et al. [163] used importance weights to develop a patient preference-based scale for assessing health outcomes in women with menorrhagia. De Bock et al. [164] assessed how well a multi-attribute value model with rank order weights predicted general practitioners' clinical decisions regarding the management of patient with suspected sinusitis. They found, inter alia, that the concordance between the model results and the actual decision was 80% in clear-cut cases of sinusitis and 50% in dubious cases [164]. Bettinger et al. [165] developed and solved a MAVT model for choosing between atypical antipsychotic treatments. Suehs and Bettinger [166] developed a MAVT model for the choice between mood-stabilizing medicines for bipolar disorder, and used a survey to obtain criteria preference weights from a large sample of specialists in psychiatric pharmacy to determine the preferred medication. After reviewing all studies included in the Cochrane review of PDA effectiveness which used computerised decision aids, we found one trials which made explicit reference to MCDA, and both referred to MAVT. In a previous RCT with the same tool, Schwartz et al. found that, within the group of women undecided about the breast cancer management option, the PDA (compared to usual care) led to an higher likelihood of reaching a decision, lower decisional conflict and higher decisional satisfaction. This study by Schwartz et al. [167] is an RCT of a 1) MAVT-based PDA for choosing between breast cancer management options in women carrying a BRCA 1/2 gene mutation versus 2) usual care. The authors found that, within the

group of women undecided about the breast cancer management option, the PDA led to a higher likelihood of reaching a decision, lower decisional conflict and higher decisional satisfaction [167]. In a subsequent study using the same PDA, Hooker et al. [168] found that women using the PDA experienced more distress than those that did not in the short term (at one month) but not in the longer term (at one year after).

8.2.2. The Simple Additive Weighting (SAW) method.

SAW is probably one of the best-known and widespread MCDA methods [169]. Suppose a choice decision problem with options A_i $(i = 1,2, \dots l)$ and criteria C_k $(k = 1,2, \dots K)$. The performance levels of the options on the criteria $x_{i,k}$ are measured quantitatively. The SAW method provides a ranking of the options using the following steps [169]:

1) Elicit weights w_k expressing the relative importance of the criteria. A number of approaches for eliciting the weights are available [169, 170], e.g. based on rating the importance of the criteria on a numerical scale or based on pairwise comparisons of the relative importance of the criteria [170];

2) Transform the performance levels $x_{i,k}$ of the options on the criteria into scores $s_{i,j}$ on a common scale. If all the criteria are increasing in preference with the performance levels, this can be achieved by dividing the performance level $x_{i,k}$ of each option by maximum performance level x_k^* [169], i.e. normalising by the score of the highest performing option:

$$s_{i,k} = \frac{x_{i,k}}{x_k^*} \qquad i = 1,2 \dots l; k = 1,2 \dots K \tag{1.4}$$

If all the criteria which are decreasing in preference with the performance levels, the appropriate transformation is [169]:

$$s_{i,k} = 1 - \frac{x_{i,k}}{x_k^*} \qquad i = 1,2 \dots l; k = 1,2 \dots K \tag{1.5}$$

If both types of criteria exist, then one appropriate transformation involves using equation (1.5) above and calculating the scores of the options on the criteria which

are decreasing in preference based on the inverse of their performance levels, i.e. [169]:

$$s_{i,k} = \frac{1/x_{i,k}}{1/x_k^*} \qquad i = 1,2 \dots l; k = 1,2 \dots K \tag{1.6}$$

3) Use a linear additive model to aggregate weights and scores into an overall score S_i for each option [169]:

$$S_i = \sum_{k=1}^{K} w_k s_{i,k} \qquad i = 1,2, \dots l \tag{1.7}$$

The scores S_i are used to rank the options from most to least preferred. Note that SAW is not designed to handle hierarchies of criteria and sub-criteria.

8.2.2.1. Annalisa, *Annalisa in Elicia* and SAW.

Annalisa is a software implementation of the "weighted sum approach" [19], i.e. a software implementation of SAW. In the software Annalisa, the scores of the options $s_{i,k}$ on the individual criteria are termed "ratings". The process of normalising the single-criterion score of the options by dividing them by the score of the highest performing option on that criterion is termed, in Annalisa, "idealisation". Idealisation is performed to ensure that, for each option A_i, each of the overall scores S_i generated in the aggregation of weights w_k and ratings $s_{i,k}$ reflects 1) the weight of each criterion and 2) the relative performances of the options on that criterion. If the options' ratings on one criterion are all very small and those on another criterion are all very high the part-worths in the overall options' scores will reflect improperly the relative performances of the options on the criteria if the ratings are not idealised (Dowie, personal communication). Annalisa software files are normally embedded using the survey software Elicia [171], which allows for customisation and personalisation of the inputs into the decision model [19]. In this thesis we refer to the joint implementation of Annalisa files and the Elicia survey software as *Annalisa in Elicia* (*ALEL*). We refer to the implementation of SAW using Annalisa as "SAW via Annalisa".

8.2.2.3. Assessment of SAW.

SAW is a long standing MCDA method. It has the advantage that it is very easy to put into practice. However, it has a disadvantage: it does not have an anchoring in axioms of rational decision making. In one variant of the SAW method explained above, where the performance levels are normalised to sum to one, it has been shown that rank reversals may occur when options are removed or added to the option set [172]. This is due to the change in the denominator of the normalisation process when the options are added/removed, which may change the scale of the resulting single-criterion scores and possibly result in rank changes [113]. To avoid this rank reversal problem, the normalisation process should be done using always the same denominator in any configuration of the decision problem [113]. The implication is that in an open system (i.e. where options can be added or removed from the set under consideration), care has to be taken when using SAW.

8.2.2.4. Application of SAW the area of diagnosis and treatment of diseases.

Based on the studies included in the review by Adunlin et al [144], there are several studies implementing SAW. Van Wijk et al [173] used SAW and TOPSIS to identify in a sample of clinicians the best first-line antihypertensive treatment. Azar [174] used SAW, TOPSIS and the Weighted Product Method (WPM) to compare the performance of several imaging techniques for diagnosing breast cancer, and found that SAW was the most robust method of the three. With respect to implementations of SAW via Annalisa in medical decision making, Masya et al. [175] demonstrated the use of a patient decision aid for the choice between alternative colorectal cancer treatment regimes. The authors elicited, in a sample of patients and in a sample of clinicians (including colorectal surgeons, medical oncologists, and radiation oncologists), individual preferences for a number of treatment outcomes using the time trade-off technique [175]. These preferences were averaged for each of the two groups of individuals and entered into an Annalisa model where the ratings of the options on the outcomes had been calculated using the best available evidence [175]. The authors found that patients and clinicians agreed on the most important outcomes to avoid, and identified the best alternative for different subgroups of individuals. Cunich et al. [176] developed and pilot-tested with a convenience sample of clinicians an Annalisa-based patient decision aid for choosing whether or

not to take a prostate screening test. They found, inter alia, that clinicians responding to a survey about the patient decision aid mostly 1) agreed with positive statements about its ease of use of the tool and 2) that the tool would be useful for discussing prostate cancer screening with their patients [176]. In the field of medical decision making, SAW via Annalisa has also been used 1) as the basis to develop a preference-sensitive measure of decision quality [21] and 2) in combination with cluster analysis to identify optimal prostate screening interventions for subgroups of individuals [177].

8.2.3. The Analytic Hierarchy Process (AHP) and the Analytic Network Process (ANP).

8.2.3.1. The Analytic Hierarchy Process (AHP).

The AHP is an MCDA method which organises "perceptions, feelings, judgments and memories into a hierarchy of forces that influence decision results" [12]. For an axiomatic representation of the AHP, see Saaty [12] and Dolan et al. [178].

The AHP decomposes the decision problem into a hierarchy of levels [12]. The simplest hierarchy is a three-level hierarchy, composed (from top to bottom) of the following elements: 1) the goal of the decision question, 2) the criteria to achieve the goal, and 3) the options [113]. Additional sub-criteria may be added under the criteria for more complex decision problems.

With a simple three-level hierarchy as illustration, the AHP operates as follows: 1) calculating local priorities (i.e. scores) for each option on each criterion, 2) calculating criteria weights, and 3) aggregating local priorities and criteria weights into a global priority (i.e. a score S_i) for each option [113]. The calculation of the local priorities of the options and of the criteria weights is based on comparing each pair of elements at the same level of the hierarchy with respect to the element immediately above it in the hierarchy [178]. To obtain the local priorities of the options, each option is compared to each other option in terms of their relative performance with respect to each criterion. To obtain the criteria weights, each criterion is compared to each other criterion in terms of their relative importance with respect to achieving the goal. In the original version of the AHP, the pairwise comparisons between elements of the hierarchy are made in terms of judgments on a

ratio 9-point scale of perceived intensity of dominance [12] which is shown in Table 1.2.

Table 1.2. AHP scale of intensity of dominance.

Intensity of dominance	Definition
1	Equal intensity of dominance
2	Weak intensity of dominance
3	Moderate intensity of dominance
4	Moderate plus intensity of dominance
5	Strong intensity of dominance
6	Strong-plus intensity of dominance
7	Very strong intensity of dominance
8	Very, very strong intensity of dominance
9	Extreme intensity of dominance

The pairwise comparisons of performance of the options with respect to each criterion are represented in a reciprocal matrix of comparative judgments. AHP allows to check the consistency of these pairwise comparisons through the consistency ratio (which adopts values between 0 and 1) [178]. The closer the consistency ratio is to 1, the closer the pairwise comparisons of performance correspond to a series of random judgments. In practice, a consistency ratio of 0.1 or less is considered acceptable [178]. If this is not the case, the pairwise comparisons of performance should to be reassessed until they are consistent.

The pairwise comparisons of performance of the options can be made, instead of using judgments on the AHP scale above, using objective data. A good example of the use of direct data in pairwise comparisons in clinical decision making is Dolan [179].

The solution for the local priorities is obtained, in the original version of the AHP, using the eigenvector method, a procedure based on matrix calculus [178]. In

contrast with the single-attribute value scores derived using the value function methods described in the previous section, which were calculated on interval scales, these local priorities are calculated on a ratio scale [180].

To calculate the criteria weights, the pairwise comparisons of their relative importance with respect to the goal are displayed in a reciprocal matrix of comparative judgments. Once their consistency is verified, the weights are calculated (using the eigenvector method). These weights are also calculated on a ratio scale. When the hierarchy contains criteria and sub-criteria, the weights of all the sub-criteria of a parent criterion reflect their relative importance with respect to their parent criterion.

The AHP uses a linear aggregation procedure to combine local priorities and criteria weights. Considering the local priorities $p_{i,j}$ of the options A_i $(i = 1,2, \dots l)$ on the criteria C_k $(k = 1,2, \dots K)$, the aggregate AHP scores S_i can be represented as:

$$S_i = \sum_{k=1}^{K} w_k p_{i,k} \qquad i = 1,2, \dots l \tag{1.5}$$

Note that these overall scores are also on a ratio scale. With such a scale, comparisons in overall scores of the type "option A_1 is *x* times better than option A_2" are meaningful. Importantly, in the original AHP the aggregation of local priorities $p_{i,k}$ across criteria is done using the distributive mode, i.e. normalising their sum to unity [113]. This may create rank reversal problems, discussed in section 8.2.3.3 below.

In an AHP hierarchy, the elements within each level of the hierarchy are assumed to be independent. For example, at the level of the criteria, two criteria are dependent if they influence each other: Ishizaka et al. [113] exemplify this type of dependence for criteria "speed" and "engine power" in a car choice decision. Dependence between options is rare [113]. Lower levels of the hierarchy are also assumed to be independent from higher levels. For example, dependence between the level of the criteria and the level of the options exists when the criteria weights depend on the options: Ishizaka et al [113] illustrate this type of dependence for a dress decision

with two criteria ("price" and "elegance") where the weight of criterion "price" changes depending on the price of a particular dress.

The AHP can be implemented with a number of software applications, e.g. Expert Choice and MakeItRational [181].

8.2.3.2. The Analytic Network Process (ANP).

The ANP is a generalisation of the AHP which does not make assumptions about the independence of the different elements of the decision problem [182], whether they be the goal, the criteria (and/or sub-criteria) or the options. In the ANP, the decision problem is not modelled as a hierarchy, but as a network of clusters, where each cluster is a collection of elements [182] (e.g. a criterion cluster or an options cluster). In a network, there are two types of dependencies: 1) dependencies between elements with a cluster, termed inner dependencies [113] and 2) dependencies between two clusters, termed outer dependencies or feedback [113].

To model inner dependencies in the criteria cluster, the ANP requires, in addition to the matrix of comparative judgments of importance of all the criteria, additional matrices of comparative judgments [113]. For example, in the case of a car choice decision problem with three criteria (price, speed and engine power), two of which are dependent (speed and engine power), three matrices are required: 1) a matrix of pairwise comparisons of the relative importance of price and speed assuming that the importance of engine power has already been assessed, 2) a matrix of pairwise comparisons of the relative importance of price and engine power assuming that the importance of speed has already been assessed, 3) a matrix of pairwise comparisons of relative importance of speed and engine power assuming that the importance of price has already been assessed. Similar procedures are used to model inner dependencies in the options cluster and outer dependencies. The different matrices of comparative matrices are combined in a supermatrix.

In the ANP, the impact of dependencies on the overall scores S_i is modelled using a Markov chain process [113].

The ANP is supported by the software Superdecisions [183].

8.2.3.3. Assessment of the AHP and the ANP.

The AHP has an axiomatic foundation (for a description of these axioms, see Saaty [184]) which focuses on the required properties of elements structured as a hierarchy in order to derive overall priorities on a ratio scale. These axioms differ from those of MAVT/MAUT in not making assumptions about the choice behaviour of a rational decision maker and have been the subject of some debate [185, 186].

The elicitation of local priorities of the options on a ratio scale separates the AHP from MAVT/MAUT, where the single-criterion values of the options are elicited on an interval scale. The use of ratio scales in the AHP have been criticised by some authors, e.g. Belton [158] and Dyer [185] because they require decision makers to (either implicitly or explicitly) establish a reference point of absolute zero overall value, a concept difficult to understand and which introduces ambiguity into the elicitation process. Harker and Vargas [180] have argued that ambiguity is a characteristic of all preference elicitation procedures, and highlight the role of the decision analyst to support decision makers in eliminating that ambiguity.

The use of the linear AHP scale (See Table 1.2) has been criticised for imposing unnatural restrictions on the decision maker's judgments, such as having an upper limit of 9 to express numerically how many more times one element dominates another element [158]. However, Harker and Vargas [180] argued that the linear AHP scale can be altered to suit an individual's need. In fact, several other scales have been proposed which increase the upper limit of this scale, such as the square root scale or the power scale [113]. Although the linear scale is the one that is most widely used, there is debate in the literature about what is the best scale [113].

As has been mentioned before, in the AHP the aggregate scores S_i of the options are calculated on a ratio scale. This means that there is an absolute zero overall score (and hence no option can have a negative overall score) and that the only meaningful comparisons in overall value between options are of the type "Option A_2 is x times better than option A_2", or, equivalently, "Option A_2 is x% better than option A_2". Belton argues that the interpretation of AHP scores is not intuitive, and that the meaning of these scores should be made clear to the decision maker at the start of the decision process [158].

In the AHP, each criterion weight should be interpreted as an "average score (average over the options under consideration) on each criterion" [158]. Salo and Hamalainen [187] have shown that if the pairwise comparisons between elements are interpreted in terms of preference differences the AHP can be interpreted as a variant of MAVT.

Perhaps the most important criticism of the AHP is the possibility of rank reversals [188]. Rank reversals of options may occur in the AHP for different reasons [113]. One reason is due to the right-left inconsistency in the eigenvector method (the method used to calculate local priorities and weights) [113]. The right-left inconsistency (or right-left asymmetry), identified by Johnson et al. [189], arises when, after replacing the pair-wise comparisons between elements (e.g. performance levels or criteria) with their inverse values the obtained ranking of those elements is not, as would logically be expected, exactly reversed [113]. Rank reversals may occur when both of the following occur: 1) the matrix of pairwise comparisons is inconsistent and 2) there are more than three elements being compared [189]. Methods alternative to the eigenvector method have been proposed to avoid the right-left inconsistency [113].

Rank reversals may also occur in the original, distributive AHP, if a copy/ near-copy of an option (respectively, an option which has the same/almost the same pairwise comparisons in terms of the performance levels than another option) is added to or subtracted from the decision model [113]. As was mentioned before, this is a violation of the Principle of Independence from Irrelevant Alternatives (PIIA) [190] and it may occur in other additive MCDA methods (such as SAW) where the performance levels are normalised [113, 172]. To avoid this type of rank reversal in the AHP, whenever options can be added to or removed from the decision model, the normalisation of the local priorities should be done using always the same denominator [113]. This is achieved dividing the local priority of each option on a particular criterion by the local priority of the best option on that criterion [113]: the resulting local priorities are termed "idealised" [191] . This is the ideal mode in AHP [113].

The ANP is a generalisation of the AHP where dependencies within and between levels of the hierarchy can be modelled. Although the ANP has the same theoretical

underpinning and the AHP [113], the procedures used to calculate the overall scores of options, based on the analysis of a supermatrix via a Markov Chain process, yield multilinear forms of aggregation of the local priorities which can be quite elaborate [192]. The interpretation of such multilinear forms is beyond the scope of this chapter. However, in terms of practical implementation, the construction of a supermatrix can be quite a labor intensive task in terms of the number of pairwise comparisons required from the decision maker.

To summarise, the use of the AHP has the advantage over MAVT/MAUT a la Keeney and Raiffa [11] that it systematically checks the consistency of the decision maker's judgments before aggregating the inputs of the decision model into overall scores. In contrast, the interpretation of these scores and of the weights is arguably less straightforward. Care must be taken in using the eigenvector method (which may lead to rank reversals if the pairwise comparisons are inconsistent). Appropriate procedures (e.g. using the ideal AHP mode) should be used when options can be added or subtracted. If there are dependencies in the model, the ANP is the appropriate approach. As with other MCDA methods, careful analysis of the results and sensitivity analysis should be used to examine the robustness of these results.

8.2.3.4. Application of AHP/ANP in the area of diagnosis and treatment of diseases.

Based on the systematic review by Adunlin et al. [144], the AHP is the most widely used method in the diagnosis and trea. The implementation of this method in clinical decision making will be illustrated with several examples, focusing in studies undertaken with patients of AHP as a tool for patient decision support. Dolan [18] pilot-tested with patients a AHP-based decision tool for the choice of colorectal screening regimen. He found that 90% of the patients were both willing and capable (where capable was understood as completing the tool in less than forty five minutes) of using the tool [18]. Dolan and Frisina [46] compared in a RCT 1) an AHP-based decision tool for choosing between alternative colorectal screening programmes with 2) an educational intervention about colorectal cancer and the screening programs. They found that the tool reduced decisional conflict and that there was no difference between patient groups in the choice of screening test [46]. Dolan [193], in a field study with primary care patients of an AHP-based decision tool for choosing

between alternative colorectal screening procedures, found using cluster analysis that there was significant variation between patients in the trade-offs they made between the decision criteria. Liberatore et al. [194] developed and successfully implemented in a general practice setting an AHP-based decision aid for deciding whether or not to undergo a prostate screening test. Carter et al [195] solved a breast cancer treatment decision problem using the AHP, the ANP and a Markov model and found that the three models gave similar results.

8.2.4. Outranking approaches Election et Choix Traduisant la Realité (ELECTRE) and Preference Ranking Organisation Method for Enriched Evaluation (PROMETHEE).

Outranking approaches such as ELECTRE [13] PROMETHEE [14] are MCDA methods which model the decision maker's preferences based on establishing binary outranking relations between options. In these methods, every pair of options is compared in order to establish one of the following situations: 1) the first option is strictly preferred to the second option, 2) the second option is strictly preferred to the first option, 3) the two options are indifferent, or 3) the two options are incomparable [13, 14]. Comparing these outranking relations between options creates a synthesising preference relational system [126] which, due to 1) the possibility that some pairwise comparisons are intransitive and 2) the possibility that some options are incomparable [126], may not be enough to lead to a clear answer to the decision question [126]. These methods use additional procedures, termed exploitation procedures, to make recommendations [126].

8.2.4.1. Election et Choix Traduisant la Realite (ELECTRE).

Different ELECTRE methods have been developed for different types of decision problems [13, 113]. For example, ELECTRE I, ELECTRE Iv and ELECTRE Is are based on selecting a subset of all the options available in such a way that only one option will be selected in the end [13]. ELECTRE II, ELECTRE III and ELECTRE IV were designed to rank from most to least preferred all the options [113]. ELECTRE-Tri-B and ELECTRE-Tri-C were developed to sort options into categories [113]. For the remainder of this section and for illustration purposes, the focus will be on ELECTRE III for ranking options.

Consider the options A_i $(i = 1,2 \dots l)$ and the criteria C_k $(k = 1,2 \dots K)$. The first stage in ELECTRE III is determining the outranking relation between the pair of options A_i and A_j $j \neq i$. The task can be summarised in the following steps [113]:

1) Construct two indices on each criterion C_k: 1.a) a partial concordance degree $c_k(A_i, A_j)$ measuring the assertion that A_i is as least as good as A_j $(j \neq i)$ on that criterion [113], and 1.b) a partial discordance degree $d_k(A_i, A_j)$ measuring the discordance with the assertion that A_i is at least as good as A_j $(j \neq i)$ on that criterion [113].

1.a) To construct the partial concordance degree for each criterion C_k, indifference and preference thresholds are defined. The indifference threshold describes numerically the largest difference between the performance levels of the criterion so that the decision maker is indifferent between two options [113]. The preference threshold describes numerically the largest difference between the performance levels of the criterion so that the decision maker prefers one option over another [113]. The partial concordance degree $c_k(A_i, A_j)$ takes values between 0 and 1 based on comparing the differences in the performance of A_i and A_j $(j \neq i)$ with the indifference and preference thresholds: $c_k(A_i, A_j) = 0$ indicates that A_i is indifferent to A_j, $c_k(A_i, A_j) = 1$ indicates that A_i is strictly preferred to A_j [113];

1.b) To construct the partial concordance degree for each criterion C_k, a veto threshold is defined. The veto threshold describes numerically the largest difference between the performance levels so that the decision maker rejects the assertion that that A_i is at least as good as A_j $(j \neq i)$ [113]. The partial discordance degree $d_k(A_i, A_j)$ takes values between 0 and 1 based on comparing the differences in the performance of A_i and A_j with the veto thresholds: $d_k(A_i, A_j) = 0$ indicates that there is no reason to reject the assertion; $d_k(A_i, A_j) = 1$ indicates that the assertion is accepted, since difference in performance between A_i and A_j exceeds the veto threshold [113];

2) Calculate the partial concordance degrees $c_k(A_i, A_j)$ and the partial discordance degrees $d_k(A_i, A_j)$ for each pair of options A_i and A_j $(j \neq i)$;

3) Assign importance weights to the criteria;

3) Calculate the global concordance degree $C(A_i, A_j)$, the weighted sum of partial concordance degrees $c_k(A_i, A_j)$ $(k = 1,2, \dots K)$ for each pair of options A_i and A_j $(j \neq i)$;

4) Calculate the global outranking degree $S(A_i, A_j)$ for each pair of options A_i and A_j $(j \neq i)$. This index (between 0 and 1) measures the extent to which option A_i outranks A_j $(j \neq i)$ [113]. It modifies the global concordance degree based on the veto effects measured by each criterion's partial discordance degree.

Once the global outranking degrees between pair of options have been established, the second stage in ELECTRE III involves using additional exploitation procedures called ascending and descending distillation procedures [113] to obtain a full ranking of the options.

Approaches have been developed to use ELECTRE for solving decision problems which are structured as a hierarchy of criteria [196].

Different ELECTRE methods are supported by different software applications [197]. ELECTRE III is supported by the software application ELECTRE III-IV [197].

8.2.4.2. Preference Ranking Organisation Method for Enriched Evaluation (PROMETHEE).

Consider the options A_i $(i = 1,2 \dots l)$ and the criteria C_k $(k = 1,2 \dots K)$. PROMETHEE methods are based on making pairwise comparisons of the difference in the performance levels of each pair of options A_i and A_j $(j \neq i)$ on each criterion C_k [14]. To undertake these comparisons, a preference function is defined for each criterion. The definition of such functions, with different possible forms (e.g. linear or Gaussian [14, 113]) depending on the decision maker's preferences and with values between 0 and 1, allows to calculate for each ordered pair of options A_i and A_j $(j \neq i)$ a preference degree $P_k(A_i, A_j)$ taking a higher value the higher the difference in performance between the options [113].

Once the criterion-specific preference degrees $P_k(A_i, A_j)$ $(k = 1,2, .. K)$ are calculated for each ordered pair of options, PROMETHEE then calculates, for each option, the positive flow (an index of preference for that option with respect to all other options [113]), the negative flow (an index of preference of all other options

with respect to that option [113]), and the net flow (the difference between the positive and negative flow). Global flows are then calculated [113] by aggregating positive/negative/net flows across criteria – this is done using criteria weights (which can be elicited in different ways, e.g. based on pairwise comparisons of relative importance or on rank ordering the criteria [113]). PROMETHEE I analyses the global positive and negative flows to obtain a partial ranking of the options, while PROMETHEE II uses the net flows to obtain a complete ranking of the options [14].

Methods exist to use PROMETHEE for solving decision problems which are structured as a hierarchy of criteria [196].

PROMETHEE methods are supported by a number of software applications, e.g. Visual PROMETHEE [198], D-Sight [199], and Smart Picker Pro [200].

8.2.4.3. Assessment of ELECTRE and PROMETHEE.

Outranking method ELECTRE and PROMETHEE have the advantage over full aggregation methods (such as MAVT/MAUT or the AHP) that, allowing for incomparability between options, they can express more flexibly the decision maker's preferences. However, some have expressed concern that the definitions and procedures used in these methods are rather arbitrary [10] and that MAVT is more transparent and provides a clearer audit trail [10]. In addition, neither PROMETHEE nor ELECTRE are free from rank reversals [113].

8.2.4.4. Application of ELECTRE and PROMETHEE in the area of diagnosis and treatment of diseases.

In the systematic review of studies presented in Adunlin et al. [144], two studies were described using ELECTRE in the area of diagnosis and treatment of diseases. One of these studies by Le Gales and Moatti [201] used ELECTRE IS to support a group of experts identify a reasonable set of screening strategies for the prevention of major hemoglobinopathies in France and choose the best alternative based on a number of criteria. The other study by Brasil Filho and Coelho [202] used ELECTRE IV for the classification of patients into different Alzheimer's disease categories. With respect to PROMETHEE, no applications in the area of diagnosis and treatment were reported in the systematic review mentioned above.

8.2.5. Technique for Order of Preference by Similarity with an Ideal Solution (TOPSIS).

TOPSIS is a reference-level MCDA method which is based on choosing options which are, at the same time, as close as possible to a positive-ideal option and as far away as possible from a negative-ideal option [129], both of which are defined by the decision maker [113]. TOPSIS is based on the following five steps [113, 129]:

1) Normalising, on each criterion, the performance levels of the options (so that they are comparable) by the sum of the squared performance levels of the options;

2) Weighting, on each criterion, the performance level of each option by the preference weight assigned to that criterion;

3) Establishing the positive-ideal and the negative-ideal options. The positive-ideal option can be 1) a hypothetical option with the best (normalised) performance level of all the available options on each criterion, 2) an absolute ideal option independent of all the available options, or 3) an option between these two [113]. Similarly, the negative-ideal option can be 1) a hypothetical option with the worst (normalised) performance levels of all the available options on each criterion, 2) an absolute anti-ideal option, or 3) an option between these two [113];

4) Calculating the distance from each option to the ideal and anti-ideal options: the standard approach is calculating, for each option, the Euclidian distance between the point defined by the vector of normalised performance levels of the option and the point defined by the vector of normalised performance levels of the positive-ideal and negative-ideal options [113]. Note that the Euclidian distance is a scalar;

5) For each option, calculate the relative closeness coefficient (a number between 0 and 1), based on dividing the distance to the anti-ideal option by the sum of the distance to the anti-ideal solution and the distance to the ideal solution [113]. The higher the relative closeness coefficient is for a particular option, the closest the option is to the ideal option and the furthest it is from the anti-ideal option.

TOPSIS can be implemented using e.g. the Tryptich Excel-based software application [203].

With respect to the assessment of TOPSIS, the main advantage is that it is a simple method to implement. However, it has been criticised because it may give illogical results [113]. As with other MCDA methods, TOPSIS is not free from rank reversals, although modifications of the original TOPSIS algorithm have been proposed to avoid this problem [204].

The only study using TOPSIS in the area of diagnosis and treatment, as reported by Adunlin et al. [144], is the study by Azar et al. [174] comparing TOPSIS, SAW and WPM in the choice of imaging technique for breast cancer diagnosis: this study showed that SAW and TOPSIS yielded similar results.

8.2.6. Conclusion.

This section has explored in detail a number of MCDA methods which 1) are long-established and/or 2) have been used in the area of diagnosis and treatment of diseases. Value function methods have the advantage that they are well-grounded on axioms of rational decision making. In this sense they are normatively more appropriate than other methods. Within these methods, MAVT/MAUT as described by Keeney and Raiffa [11] may be difficult to implement. For this reason, simpler approaches like SMARTS and SMARTER have been proposed to facilitate the MCDA process. Another approach, MACBETH, may be impractical when the decision involves many options or many criteria, as it requires to make many pairwise comparisons. This applies to the AHP which, although axiomatically anchored, does not make assumptions about the rational behaviour of the decision maker. The AHP has to be used with care, as in decision situations where the options are not fixed it can give rise to anomalous results (rank reversals) when options are added or removed. SAW, a very easy methods to implement, also has to be used with care for the same reason. TOPSIS is also easy to implement, but it may give rise to illogical results. ELECTRE and PROMETHEE allow to express more realistically the preferences of decision makers than full aggregation methods such as MAVT/MAUT, AHP and SAW, but they are not very transparent and may also lead to rank reversals.

Based on a systematic review of studies [144] of the application of these methods in the area of diagnosis and treatment, the most used method in patient decision support is the AHP. The AHP has been described as acceptable to users and has successfully

been applied in a general practice setting. SAW has been successfully used by clinicians to choose between antihypertensive treatments. SAW via Annalisa has been described by clinicians as easy to use and useful to provide clinical decision support. Value function methods have been used by health practitioners to make decisions, and also to develop preference-based health outcome scales. It has been shown that patients are willing to use the weight elicitation procedures of SMART and SMARTS. ELECTRE has been used to choose between alternative screening programmes for prevention of major hemoglobinopathies, but not at patient level and to help clinicians diagnose Alzheimer's disease patients. No information was reported in Adunlin et al [144] about uses of TOPSIS or PROMETHEE with patients or clinicians.

8.3. Choosing an MCDA method.

As has been shown in the previous section, different MCDA methods differ not only in their axiomatic basis, but also in the specific implementation of the MCDA process. Choosing a MCDA method is not a straightforward task [205]. Different approaches which have been proposed include 1) choice rules and algorithms (including expert systems) based on the information required by the model, 2) field experiments with subjects exploring their reaction to different methods in terms of specific criteria, and 3) comparison experiments in the form of mathematical simulations [206].

Studies based on choice rules and algorithms are more useful for eliminating than for choosing one particular MCDA method [206]. Field studies are useful to compare MCDA methods in terms of user reactions, but they have limitations (e.g. they tend to have small sample sizes, the way information is elicited may influence the results more than the method used, learning effects may bias the outcomes) [206]. Existing simulation experiments exploring the operational aspects of different MCDA methods have been undertaken on a variety of these methods, and summarising this literature is beyond the scope of this review. Three of these studies are commented on here because they involve several of the methods described in the previous sections. Buede and Maxwell [207] performed a number of simulations to compare MAVT (which was used as a benchmark because it is immune to rank reversals) with several methods known to show rank reversals, including the AHP and TOPSIS. The

results showed that while the rank disagreements between MAVT and TOPSIS were likely, they were unlikely between MAVT and the AHP [207]. Zanakis et al. [206] performed simulations comparing the performance of several MCDA methods including SAW (used as a benchmark), Multiplicative Exponent Weighting (MEW), four versions of the AHP, ELECTRE and TOPSIS. They found, inter alia, that 1) the AHP versions performed closer to SAW than the other methods and that 2) when the number of options increase, the methods produce similar final weights, dissimilar rankings, and more rank reversals [206]. Salminen et al. [208] performed simulations comparing SMART, ELECTRE III and PROMETHEE I and II. They found that in many situations the three methods performed similarly, but that the preferred options can vary greatly between methods in specific situations [208].

In a review of empirical studies comparing the results of different MCDA methods when applied to the same decision problem, Mysiak [209] reported that:

- In terms of the perceived usefulness of methods by different users: methods which rated highly on ease of use often rated poorly in inspiring confidence in the results; harder methods gave more similar results across users; simpler methods were generally reported as less appropriate;

- The evidence regarding the consistency of results when the same individual uses different MCDA methods is contradictory;

- When the same MCDA method is used by different individuals, the results can vary greatly due to interpersonal differences in preferences and in experience;

- Studies evaluating the difference in the consistency of results when both 1) the same individual is exposed to different methods and 2) different individuals are exposed to the same method found no evidence to that effect;

The experimental validation of MCDA methods is problematic, as there is no agreed definition of validity [210]. In the light of this observation, is there a best MCDA method? Several authors suggest that the “right” method will depend on the characteristics of the specific decision situation [209, 211]. Salminen [208] and Mysiak [209] both recommend to use more than one MCDA method to solve a decision problem. Buede and Maxwell [207] argue that structuring the decision problem (in terms of the selection of options and criteria) and adequately eliciting of

the inputs into the decision model are as important as the choice of method. Philips argues that decision models are requisite when their "form and content are sufficient to solve a particular problem" [212]. Such decision models are constructed in an interactive process of consultation with a decision analyst: extensive sensitivity analysis is used to explore the results and reshape (if necessary) the model until no more intuitions arise about the problem [212]. The focus of these models is on analysis, and in this sense they are not necessarily prescriptive [212]. In order to assess the validity of requisite decision models, multi-criteria requisite evaluation models can be developed, where the evaluation criteria may be objective or relatively subjective [212].

8.4. Software applications for implementing MCDA methods.

An ample variety of software applications are available to implement MCDA methods. Weistroffer et al. [213] provides an overview of these tools. The 2014 Decision Analysis Software Survey by OR/MS Today [214] includes a number applications implementing MCDA, and also includes detailed vendor-provided information about their features. The International Society of Multi-Criteria Decision Making provides, on its website [215], links to the websites of a number of tool developers/vendors.

8.4.1. Software applications: types and characteristics.

Some applications are MCDA-method-specific, such as the aforementioned ELECTRE family of applications [197], Visual PROMETHEE [198], M-MACBETH [154] and *Annalisa in Elicia* (*ALEL*) [20]. Others allow for the combination of features of different MCDA methods. A good example of this second type is Web-HIPRE [216]. Web-HIPRE uses a linear additive MAVT model but the weights can be elicited using a number of techniques, e.g. direct rating, swing weighting, or pairwise comparisons of importance as in the AHP [217]. HiView 3 [15] allows to use MAVT (with swing weights) but it also permits 1) the use of MACBETH to evaluate a whole model and 2) the use of MACBETH only for weight elicitation. Although many software applications implementing MCDA are generic, some have been designed for specific fields. For example, PurE2 [218] was developed to solve multi-criteria decision problems in the field of urban pollution. MCDA-Res [219] was designed to solve multi-criteria renewable energy decision problems.

Applications developed in academic institutions are often available for free (or for a small fee), but commercial vendors (some of which give educational discounts) can charge hundreds or thousands of dollars [213]. Mustajoki and Marttunen [220] provide relatively recent cost information for commonly used MCDA applications.

The purpose of this section is to review different generic MCDA software applications for solving a particular choice decision problem. Because of the large number of existing applications implementing MCDA, a decision was made to focus on a reasonable subset. In a review of these applications and their features, Mustajoki and Marttunen [220] identified twenty-four applications that have "been actively used or that have achieved some status among the practitioners and MCDA community (which can be seen as an indication of the software offering such features that make it worth using)" [220]. The set of applications identified by these authors was used as a starting point for the review presented here. For inclusion in this review, the following software applications were eliminated: 1) those that were not generic; 2) in addition, those that were not based on the MCDA methods described in sections 8.1. and 8.2. This resulted the following applications for review: Analytica [221], Criterium Decision Plus [222], D-Sight, GMAA, HiView 3, Logical Decisions, M-MACBETH, MakeItRational, OnBalance [223], Smart Decisions (previously Promax) [224], V.I.S.A Decisions, and Web-HIPRE [216, 217]. Expert Choice was added to the set since it is a well-known AHP software and has been used in medical decision making. *ALEL* was added to the set because it has been used previously in medical decision making. Studies comparing MCDA applications were identified from the literature, i.e. French and Xu [225], Mustajoki and Marttunen [220], and Baizyldayeva et al. [226]. The results of these studies and and the OR/MS survey [214], complemented by any user guides available for each package and trial testing of the software (where possible) were used to assess the applications included in this review. The focus of the review is on how each application addresses the MCDA process. Information is also provided on any available features within the packages that provide users with support through the decision making process, e.g. in terms of 1) tab panels distinguishing each step of the decision-making process or 2) step-by-step guidance through decision-making [220]. Information is also provided on any explicit group decision support available with each package. The results of the review are summarised in five tables, respectively

commenting for each application on: methods supported, decision support features (user manuals and help menus not included, as they are common software features), and group decision support (Table 1.3); problem structuring (Table 1.4); scoring options on the individual criteria (Table 1.5); weighting criteria (Table 1.6); analysis of results and sensitivity analysis (Table 1.7).

Table 1.3. Software applications for MCDA: methods implemented, decision support features and group decision support.

Software	**Description of method(s) implemented, decision support features, and explicit support for group decision making**
Analytica	MAVT with additive MAUF. No decision support features. Explicit support for group decision making not provided.
ALEL	SAW. Decision support can be added through the Elicia functionality, which allows to tailor the steps of the decision process via a sequence of screens. *ALEL* allows for different users to work on the same model remotely and supports decentralised elicitation of information.
Criterium Decision Plus	SMART and AHP. No decision support features. Explicit support for group decision making not provided
D-Sight	MAVT/MAUT with additive MAVF/MAUF and PROMETHEE. Provides decision support in tabs separating the elements of the decision-making process. D-sight allows for different users to work on the same model remotely and supports decentralised elicitation of information.
Expert Choice (version 11.5[2])	AHP. No decision support. Group decision making is supported by creating models that can be worked on remotely by several members of a group. In group mode, the judgments of the different users can be combined.
GMAA	MAVT/MAUT with additive MAVF/MAUF. Allows for imprecision in value elicitation. No decision support. Explicit support for group decision making is not provided.

[2] The author of this thesis did not have access to a more recent version of this software

Table 1.3 (cont.). Software applications for MCDA: methods implemented, decision support features and group decision support.

Software	**Description of method(s) implemented, decision support features, and explicit support for group decision making**
HiView 3	MAVT with additive MAVF. MACBETH supported. No decision support. Explicit support for group decision making is not provided.
Logical Decisions	MAVT with additive MAVF, AHP, and AHP as a value function approach. The "Logical Decisions Facilitator" provides information about the steps of the decision-making process in tab panels. Explicit support for group decision making is provided in Logical Decisions for Groups, but not in the standard version of the software.
M-MACBETH	MACBETH. No decision support. Explicit support for group decision making is not provided
MakeItRational	AHP. Provides decision support in tabs separating the elements of the decision-making process. Group decision making is supported by creating models that can be worked on remotely by several members of a group. In group mode, the judgments of the different users can be combined.
OnBalance	MAVT with linear additive MAVF. No decision support. Explicit support for group decision making is not provided.
Smart Decisions	MAVT with linear additive MAVF. Provides decision support in tabs separating the steps of the decision-making process. According to the OR/MS Survey, explicit support for group decision making is provided in Smart Decisions Gold [214].
V.I.S.A. Decisions	MAVT with linear additive MAVF. Provides decision support in both tab panels and in a step-by-step guide of the decision process. Explicit support for group decision making is not provided.
Web-HIPRE	MAVT with linear additive MAVF and AHP as a value function method. No decision support. Group models can be created combining the weights of different users.

Table 1.4. Software applications for MCDA: problem structuring.

Software	**Description**
Analytica	Problems structured as influence diagrams. Hierarchies can be modelled and displayed. Easy to add and drag elements across the screen. Interfaces can be created for users to enter information without having to view the entire diagram
ALEL	Does not have an interface for structuring the decision problem. Does not allow for criteria hierarchies.
Criterium Decision Plus	Criteria hierarchies can be modelled and displayed. "Brainstorm window" to structure decision problems, where elements can easily be added and dragged around. The resulting model in a "Brainstorm window" can automatically be transformed into a hierarchy. Hierarchies are represented in the "Hierarchy window".
D-Sight	Criteria hierarchies can be modelled and displayed. Criteria can be added and dragged around easily.
Expert Choice	Criteria hierarchies can be modelled and displayed. Two different interfaces: the "Cluster view" interface and the "Treeview" interface. Both are flexible in terms of arranging criteria. Additionally, the "ProCon pane" allows to set pros and cons of options and then convert them into criteria.
GMAA	Criteria hierarchies can be modelled and they are displayed from left to right. The problem structuring interface allows to add criteria but it does not allow to drag criteria around the screen

Table 1.4 (cont.). Software applications for MCDA: problem structuring.

Software	**Description**
HiView 3	Criteria hierarchies can be modelled and displayed vertically or horizontally. Criteria can be added and dragged around easily
Logical Decisions	Criteria hierarchies can be modelled and are displayed horizontally or vertically. "Brainstorming window" to structure the hierarchy, criteria can easily be added and be dragged around the screen.
M-MACBETH	Criteria hierarchies can be modelled and displayed from left to right. Hierarchies are built using "nodes", of which there are two types: "non-criteria nodes" and "criteria nodes". The first type can be used to structure the decision model in a hierarchy, but contain no information. The second time are the nodes on which the model is assessed. Nodes can be added but cannot be dragged around the screen.
MakeItRational	Criteria hierarchies can be modelled. In the online version, criteria can be added but not dragged around, and the hierarchy cannot be displayed graphically. In the desktop version the hierarchy can be displayed graphically but the criteria cannot be dragged around the screen
OnBalance	Criteria hierarchies can be modelled, and are displayed from top to bottom. Criteria can be added but not dragged around on the screen. Criteria are divided into "benefit" and "cost" criteria

Table 1.4 (cont.). Software applications for MCDA: problem structuring.

Software	**Description**
Smart Decisions	Criteria hierarchies can be modelled and displayed, criteria can be added but not dragged around
V.I.S.A. Decisions	Criteria hierarchies can be modelled and displayed. Criteria can be added and dragged around the screen easily.
Web-HIPRE	Criteria hierarchies can be modelled and displayed. Criteria can be dragged around the screen easily

Table 1.5. Software applications for MCDA: scoring options on the individual criteria

Software	**Description**
Analytica	Allows for visual assessment of single-criterion value scores for each option. Not possible from surveys or user guide to identify the exact value elicitation procedures available in the software
ALEL	Single-criterion scores for each option assessed directly either in text or on a sliding bar

Table 1.5 (cont.). Software applications for MCDA: scoring options on the individual criteria

Software	**Description**
Criterium Decision Plus	Allows for visual assessment of single-criterion value scores for each option. Not possible from surveys to identify the exact value elicitation procedures available in the software. From the user manual, information is provided about the AHP elicitation of local priorities. For calculating local priorities, pairwise comparisons of relative performance are done on a "rating" screen which allows for numerical, verbal or graphical comparisons between options.
D-Sight	According to Mustajoki and Marttunen [220], the software does not allow for visually eliciting single-criterion value scores. This information could not be verified as this author did not have access to the user manual.
Expert Choice	Single-criterion local priorities for each option are elicited using pairwise comparisons of relative performance with three possible formats: verbal, numerical, or with sliding bars. Local priorities can be entered directly if data is available. Inconsistencies are displayed and suggestions made for their correction. The software allows to construct value functions, a departure from AHP

Table 1.5 (cont.). Software applications for MCDA: scoring options on the individual criteria

Software	**Description**
GMAA	Support in the elicitation of single-criterion value scores is provided using visual aids. Option value scores can be elicited directly using a vertical value scale (i.e. thermometer). For discrete performance levels, discrete value functions can be built entering directly the value scores to construct a graph. For continuous performance levels, linear or piecewise linear value functions can be built on a graph. Single-attribute utility functions ca be built using certainty equivalence methods and probability equivalence methods with visual aids. Imprecision can be built into the functions.
HiView 3	Support in the elicitation of value scores is provided using visual aids. Option value scores can be elicited directly using vertical value scales (i.e. thermometers). For discrete performance levels (e.g. verbal levels), discrete value functions can be built using vertical value scales. For continuous performance levels, single-criterion linear value functions are the default in the software, but piecewise linear value functions can be constructed using interactive graphs. Logarithmic value functions can be used to represent uncertainty as a criterion. In addition, MACBETH scales can be built using a matrix of pairwise comparisons of difference in attractiveness (the M-MACBETH interface)

Table 1.5 (cont.). Software applications for MCDA: scoring options on the individual criteria

Software	**Description**
Logical Decisions	Support in the elicitation of single-criterion value scores is provided using visual aids. Option value scores can be elicited directly using a screen with sliding bars. The balanced beam method (based on equally preferred bundles of items) can also be used to elicit option value scores. For continuous performance levels, linear, piecewise and exponential value functions can be assigned using an interactive screen. The mid-value splitting technique can also be used to elicit option value scores. Option utility scores can be assessed using lotteries. For eliciting local priorities using the AHP, a matrix of numerical pairwise comparisons is available, including consistency checks. The local priorities can be normalised in several ways: 1) assigning the highest performing option a priority of 1 and the lowest performing option a priority of 0; normalising priorities so they add to 1 (the distributive AHP); normalising by the priority of the highest performing option, i.e. idealising (the ideal AHP)
M-MACBETH	Single-criterion value scores are elicited in a matrix of pairwise comparisons of difference in performance, with inconsistencies displayed and suggestions made for their correction. Several options are available for these pairwise comparisons: 1) comparing the local options, 2) comparing the local options + two reference options, 3) comparing qualitative performance levels, and 4) comparing quantitative performance levels. Allowance is made for adjusting the resulting scale graphically.

Table 1.5 (cont.). Software applications for MCDA: scoring options on the individual criteria

Software	**Description**
MakeItRational	Single-criterion local priorities for each option are elicited using pairwise comparisons of relative performance with a visual aid, with inconsistencies displayed and suggestions made for their correction.
OnBalance	Support in the elicitation of option value scores is available with visual aids. Option value scores can be elicited directly using vertical value scales (i.e. thermometers). For discrete performance levels (e.g. verbal levels), discrete value functions can be built using histograms. For continuous performance levels, single-criterion linear value functions can be constructed. Piecewise linear value functions can be elicited using interactive graphs. The software allows to visualise the single-criterion value scores and the performance levels in visual scales side by side.
Smart Decisions	Support in the elicitation of option value scores is available with visual aids. Linear single-criterion value functions are the default option, but piecewise linear value functions can also be created interactively.
V.I.S.A. Decisions	Support in the elicitation of value scores is provided using visual aids. Option value scores can be elicited directly using vertical value scales (i.e. thermometers). For discrete performance levels (e.g. verbal levels), discrete value functions can be built using vertical value scales. For continuous performance levels, single-criterion linear value functions and piecewise linear value functions (the latter can be elicited using interactive graphs)
Web-HIPRE	Single-criterion value scores for each option can be entered directly. Single-criterion value functions (linear, piecewise or exponential) can be elicited. Value scores can also be calculated making pairwise comparisons of difference in attractiveness (the AHP as a value function approach)

Table 1.6. Software applications for MCDA: weight elicitation.

Software	**Description**
Analytica	From Mustajoki et al. [220] criteria weights can be elicited in different ways, including swing weights. The visual aids for weight elicitation were not found in the software user guide
ALEL	Criteria weights can be entered as numbers or can be elicited on sliding bars
Criterium Decision Plus	From Mustajoki et al. [220], different weight elicitation procedures are possible, including swing weights and AHP weights. From the user manual, the only elicitation procedure that could be observed was that of pairwise comparison of relative importance of criteria (AHP weights). For calculating criteria weights, pairwise comparisons of relative importance are done on a "rating" screen which allows for numerical, verbal or graphical comparisons between criteria.
D-Sight	According to Mustajoki and Marttunen [220] weights can be elicited visually. The visual aids for weight elicitation could not be found.
Expert Choice	Criteria weights are elicited using pairwise comparisons of relative importance with three possible formats: verbal, numerical, or with sliding bars. Criteria weights can also be entered directly. Inconsistencies are highlighted by the software, which provides suggestions for their correction

Table 1.6 (cont.). Software applications for MCDA: weight elicitation.

Software	**Description**
GMAA	Weight elicitation is undertaken using trade-off methods or directly assigning weights (N.B. weight intervals) to each of the criteria.
HiView 3	Criteria weights can be entered directly and elicited using swings with a visual aid. Criteria weights can also be assessed using comparisons of difference in attractiveness with the M-MACBETH functionality.
Logical Decisions	Criteria weights can be elicited with visual aids in a number of ways, e.g. direct rating, the trade-off method, swing weights, rank ordering, pairwise comparisons of criteria importance (for the AHP, this approach checks consistency in judgments).
M-MACBETH	Criteria weights are elicited in a matrix of pairwise comparisons of difference in attractiveness, with inconsistencies displayed and suggestions made for their correction. The resulting criteria weights can be displayed as bar charts and adjusted within consistent levels.
MakeItRational	Criteria weights are elicited using pairwise comparisons of relative performance with a visual aid.
OnBalance	Criteria weights are elicited using trade-off methods or swing weights with visual aids

Table 1.6 (cont.). Software applications for MCDA: weight elicitation.

Software	**Description**
Smart Decisions	Criteria weights can entered directly in numerical form or elicited using swing weights with a visual aid or using numerical pairwise comparisons (AHP as a value function approach)
V.I.S.A. Decisions	Criteria are elicited using swings with a visual aid
Web-HIPRE	Criteria weights can be elicited using direct rating, swings, ranking of options + point distribution (SMART), rank ordering (SMARTER), and pairwise comparisons of difference in attractiveness (AHP as a value function approach)

Table 1.7. Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
Analytica	Overall option scores are visualised numerically, although they can also be viewed graphically. Sensitivity analysis on changes in individual weights can be undertaken for each overall score. Two-way sensitivity analysis can also be undertaken. Tornado diagrams can be generated to explore the sensitivity of overall results to changes in several weights

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
ALEL	Overall option scores are visualised as a horizontal bar graph at the top of the screen, accompanied by the numerical values. The package allows for overall results to be idealised (normalised by the highest overall score) or distributed (normalised to add to 1). Stacked bars showing the contribution of each criterion to the overall scores. Although *ALEL* supports no sensitivity analysis screens separate from the main screen. The sensitivity of the overall option scores to changes in weights (or in single-criterion ratings) can be visualised directly in the main screen: as the weights (or ratings) are changed, the overall scores of the options change accordingly
Criterium Decision Plus	From the user manual, overall option scores are presented visually on a horizontal bar graph, accompanied by the numerical values. Stacked bar graphs can be displayed to visualise the contribution of each criterion to the overall score of each option. The "sensitivity by weights" screen plots each criterion weight against the overall scores of the options and allows for assessing interactively how much a criterion weight needs to change for a change in the overall ranking of the options.
D-Sight	From Mustajoki and Marttunen [220], visual graphs are provided and there is a sensitivity analysis functionality. The implementation of this feature could not be verified due to lack of access to the software or to its user manual

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
Expert Choice	Overall option scores are visualised in a horizontal bar graph with accompanying overall priorities, and can be presented using the ideal or distributive AHP mode. There are five sensitivity analysis screens: 1) the "dynamic" screen allows the user to assess in real time the impact on the overall scores of the options of changing the weights of the top-level criteria using bar graphs; 2) the "performance" screen allows for a similar assessment while visualising the contribution of each option to each top-level weight; 3) the "gradient" screen plots each criterion weight against the overall scores of the options and allows for assessing interactively how much a criterion weight needs to change for a change in the overall ranking of the options; 4) the "two-dimensional" screen (not interactive) plots the priorities of the options for any pair of criteria, and allows to inspect situations of dominance and key trade-offs; 5) the "head to head" screen allows to compare two options overall and for each top-level criterion.
GMAA	Results are presented in a horizontal bar graph with accompanying numerical values. Both mean overall scores and any upper or lower bounds (due to the effect of imprecision in value elicitation) are presented. A stacked bar graph can be used to visualise the contribution of each criterion to the overall scores. A vertical bar graph allows to visualise the contribution of each criterion to the overall score of each option. The "compare optionss" graph allows to visualise differences between options along the criteria and overall. In terms of sensitivity analysis, weights can be changed to explore their effect on overall option scores, but this feature requires moving between screens (it is not interactive). The software has a feature to assess dominated and potentially optimal options. It also has a feature for sensitivity analysis of weights using Monte Carlo simulations.

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
HiView 3	Overall option scores are visualised in a vertical bar graph with accompanying numerical values and accompanying weights of the top-level criteria. The user can switch to visualising in stacked bars the contribution of each criterion to the overall score of each option. The impact of changing criteria weights on the overall numerical scores can be assessed interactively using sliding bars. The "map" screen plots 1) the weighted scores of the options against criteria two at a time and 2) the efficient frontier, allowing for assessment of dominance and key trade-offs. The "sorts" screen compares options two at a time with respect to each of the criteria. Different sensitivity analysis screens are available: 1) the interactive "sensitivity up" screen allows to visually assess how changes in each criterion weight impacts on the overall scores and ranking of the options; 3) the "sensitivity down" tool identifies with colour codes which criteria weights are more or less sensitive (red indicates that a criterion would have to increase/ decrease less than 5% for the highest ranked option to change, orange indicates that a criterion would have to increase/decrease between 5% and 10% for the highest ranked option to change, green indicates that a criterion would have to change by more than 15% for the highest ranked option to change).

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
Logical Decisions	Overall option scores are visualised in horizontal bar charts with accompanying numerical values. The contribution of each criterion to the overall scores can be visualised in stacked bars. A "scatter diagram" screen allows to visualise the weighted scores of each option on any two criteria. There are several sensitivity analysis screens, e.g.: 1) An interactive "dynamic sensitivity" screen allows to visualise with sliding bars the impact of changing weights on the overall scores, 2) a "compare options" screen allows to visualise in bar charts the differences in overall scores and single-criterion scores for every pair of options, 3) a "sensitivity graph" which allows to graphically assess the impact of changes in individual weights on the overall score of the options.
M-MACBETH	Overall option scores can be visualised on a scale. An "option profiles" screen allows to visualise each option's single-criterion scores in relation to the lower and upper reference levels. "Difference profile" screens allow to visualise the difference in single-criterion value scores between pairs of options. "XY" maps allow to compare the single-criterion values of the options along pairs of criteria. There are several types of sensitivity analysis: 1) the "sensitivity analysis on weight" screen allows to visualise the effect of changing individual weights on the ranking of the options; 2) the "overall thermometer" feature allows to visualise on a scale the impact on the overall option scores of changing a) the value score of an option on a particular criterion or b) the weight of a criterion. The "robustness analysis" screen allows to identify dominant and additive dominant options.

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
MakeItRational	Overall option scores are visualised in a horizontal bar graph, and can be presented using the ideal or distributive AHP mode. Stacked bars can be used to visualise the contribution of each criterion to each overall option score. The options can be compared across criteria using a radar chart. The sensitivity analysis screen allows to visualise how changes in individual criteria weights affect the overall scores and the ranking of options
OnBalance	Overall option scores are visualised on horizontal bar graphs with or without their associated numerical scores. Stacked bars can be used to visualise the contribution of each criterion to each overall option score. The "map" window plots option scores on two criteria to assess dominance and key trade-offs. A "sensitivity on node" screen allows to graphically assess the impact of changes in individual weights on the overall score of the options. A "test robustness" window allows to compare options two at a time to visualise how much better (or worse) one option is than the other on each criterion.

Table 1.7 (cont.). Software applications for MCDA: analysis of results and sensitivity analysis.

Software	**Description**
Smart Decisions	Overall option scores are visualised on horizontal bar graphs with their associated numerical scores. The contribution of each criterion to the overall scores can be visualised in stacked bars. A "strengths and weaknesses" screen allows to see the differences between options on each of the criteria. A "scatter" screen displays scores of options on two criteria at a time to assess situations of dominance and key trade-offs. Weights of individual criteria can be plotted against overall option scores to assess when changes in these weights alter the ranking of options. It is possible to also assess for each criterion the range of weights over which the ranking of options will change.
V.I.S.A. Decisions	Overall option scores are visualised in a vertical bar graph. "Score profiles" allows to compare the weighted scores of the options for each criterion in a single graph. The "dominance" table allows to compare the options in terms of dominance. An "XY" screen allows to compare option scores two criteria at a time to assess dominance situations and key trade-offs. A "weight sensitivity" screen allows to visualise how changes in individual criteria weights affect the overall scores and the ranking of options
Web-HIPRE	Overall option scores are visualised in vertical bar charts with accompanying numerical values. The contribution of the criteria/options to each overall score can also be visualised. The sensitivity analysis screen allows to view how changes on a criterion weight impacts on the overall scores and on the ranking of the options

8.4.2. Assessment of software applications for MCDA.

In order to choose a software application to implement the MCDA process it is important to consider the context of the decision and the process of the decision [225]. Decision context characteristics pertain to 1) the type of decision problem (e.g. whether it is more or less structured, the number of criteria/options to be considered), 2) the social context (e.g. who is/are the decision maker/s and what are his/her/their responsibilities), and 3) the cognitive factors of the decision maker/s [225]. Decision process characteristics include 1) whether the final decision maker is one individual or several individuals with potentially conflicting goals and worldviews, 2) the time constraints for making the decision, 3) whether or not the results of the analysis needs to be communicated to stakeholders not involved in decision-making, and 4) how the analysis will be conducted (e.g. working through the problem with an analyst, as is the case in a decision conference [227], by the decision maker him/herself) [225].

This thesis focuses on the development and implementation of MCDA-based PDAs for complex decisions by clinicians in the environment of routine clinical practice in Spanish NHS hospitals. Given this, the decision context and the decision process have some general characteristics. With respect to the decision context:

1) The decision problem is relatively structured, i.e. the options are limited to the available health care options (e.g. treatments) for the condition at hand and the decision criteria are limited to the consequences of the options which are relevant for the patient;

2) The decision maker is a patient making a complex decision with inputs from his/her clinician;

3) The cognitive features of different patients will differ, but it should not be assumed that an average patient has high levels of health literacy.

With respect to the decision process:

1) The final decision maker is the patient;

2) Time and other resource constraints exist both in the development and in the implementation of MCDA-based PDAs by clinicians in routine clinical practice;

3) The development of an MCDA-based PDAs for a complex decision by clinicians in routine clinical practice is likely to involve a team of health practitioners;

3) The main stakeholders to which the decision needs to be communicated are mainly the patient's relatives and the patient's health practitioners;

4) In general, the presence of an analyst to support the development and delivery to patients of MCDA-based PDAs in routine clinical practice should not be assumed.

The software applications described in this review share one obvious characteristic: they all implement MCDA. In this sense, their use has the potential to improve the quality of the decision process and the quality of the decision made, two key effectiveness measures of PDAs. They differ in a number of aspects, including:

1) In the MCDA method implemented. The choice of software application will determine the MCDA method implemented unless the application supports different methods (e.g. Logical Decisions or Web-HIPRE). One advantage of supporting different methods is that the end user can choose the one which best suits his/her cognitive style. However, some applications (e.g. Web-HIPRE) which allow the user to mix methods may result in incompatibilities with the theoretical underpinnings of the methods [225]. Applications which implement prescriptive methods (such as HiView, V.I.S.A. decisions, or Logical Decisions) have a clear advantage in terms of best practice decision-making over those that do not (such as *ALEL*). Elicitation procedures used by different methods differ in complexity. Logical Decisions, for example, supports the elicitation of weights using trade-offs, which can be cognitively challenging. *Expert Choice* and M-MACBETH are based on pairwise comparisons, which can be tedious with many options/criteria. *ALEL* has an advantage here in requiring little elicitation effort;

2) In the decision support features. In the context of the development and implementation of MCDA-based patient decision aids in routine clinical practice, decision support is important as patients/clinicians are usually not familiarised with the MCDA process. Applications which separate the steps of decision-making with tabs (e.g. MakeITRational, D-Sight) have an advantage over those that do not (e.g. *Expert Choice*, Analytica). Some applications provide step-by-step guidance through the decision-making process. This feature is in-built in V.I.S.A. Decisions. *ALEL* can incorporate this feature through the Elicia functionality, but it has to be built by the

tool developer as a sequence of screens. A more sophisticated approach would be the use of natural language generation techniques to incorporate automated explanations into the software, as was proposed by Papamichail and French [228];

3) In the type of group decision support offered. In the development stage of the MCDA-based PDA, a fit-for-purpose software application will facilitate the interaction of those involved in developing the tool, be they only health practitioners or health practitioners and patients. This interaction can take several forms. For example, a meeting (which may be a decision conference with an analyst [227]) to agree on the structure of the decision model. Or remote interaction to populate the performance matrix interface. In the case of meetings, projection of the application interface on a screen will facilitate interaction. Here, fit-for-purpose applications should have screens clear of distractions and with text and plots easy to read [225]. This author does not know whether any of the applications reviewed have been designed for projection, but in the experience of this author with trial versions of the software MakeITRAtional, Smart Decisions and V.I.S.A. Decisions stand out as clear of distractions and with texts and plots easy to read. *ALEL* also stands out in this sense, but it has the important disadvantage that it does not allow for building hierarchies of criteria interactively. HiView, designed to support decision conferencing [225], is a good alternative if an analyst is involved. Applications supporting remote interaction include *ALEL*, *Expert Choice*, D-Sight, and Logical Decisions for Groups. Applications that do support remote interaction include GMAA, OnBalance, and Criterium Decision Plus. Group interaction, however, might also be required in the implementation of MCDA-based PDAs in routine clinical practice. In this case, interaction will be between the patient and his/her health care practitioner. Applications which facilitate this interaction have minimal distractions and easy-to-access information, e.g. pop-ups when dragging the mouse over a criterion or a score. *ALEL* stands out in this regard.

4) In the functionality for structuring the decision problem. In the development of patient decision aids by clinicians, the availability of easy to use interfaces to structure a hierarchy of criteria (as is the case with e.g. Criterium Decision Plus, HiView, V.I.S.A. Decisions, and *Expert Choice*) is an advantage. Here, software applications like Web-HIPRE which 1) display the options in the hierarchy and which 2) display lines connecting all the elements of the hierarchy are at a

disadvantage. Applications like *ALEL* which do not support hierarchical structuring of decision problems are not useful for structuring the decision problem;

5) In the visual aids for the elicitation of inputs from the decision maker. In delivering PDAs to patients, visual aids should be simple and easy to access. *ALEL* ranks highly in this respect: all the inputs are elicited using sliding bars and all the information is contained in one screen. V.I.S.A. Decisions also ranks highly, its elicitation screens are uncluttered and the user can access them with a minimal number of clicks. In this sense, Analytica and GMAA rank poorly: they require substantial numerical input and navigation between screens is not designed for non-initiated users. In OnBalance, it is not easy to navigate to the elicitation screens, but the visual aids for input elicitation are very clear. HiView and *Expert Choice* also have very clear elicitation screens, although navigation to these screens for non-initiated users is not easy without support;

6) Analysis of results and sensitivity analysis. In delivering MCDA-based PDAs to patients, results should be easy to visualise and understand. *ALEL* is one of the best applications here: the graph chart showing the overall scores of the options is always visible, it occupies one third of the computer screen and the bar for the preferred option is highlighted in a different colour than the bars of the other options. Here, the availability of screens comparing alternatives in pairs to assess how much better one option is to another option overall and on individual criteria is useful. Several applications offer this possibility, e.g. *Expert Choice* and Logical Decisions. The importance of sensitivity analysis in MCDA has been highlighted by e.g. Dodgson at al [10] and Phillips [212], and methods for expert users of software have been explored e.g. by Hodgkin et al. [229]. In the context of this thesis, sensitivity analysis should be informative while supporting different levels of health literacy in different patients. Here, screens like the "dynamic" sensitivity screen in *Expert Choice*, which shows with interactive slide bars how changes in weights affect overall scores in real time, are very valuable. HiView is unique in providing easy to visualise and informative sensitivity analysis with its "sensitivity down" screen, which identifies with colour codes which criteria weights are more or less sensitive. A similar feature to "sensitivity down" is available in Smart Decisions, but the sensitivity of criteria weights is represented numerically, and hence harder to interpret than if it was colour-coded.

How should clinicians choose between alternative MCDA software applications for developing and implementing PDAs for complex decisions in routine clinical practice? Jadhav and Sonar [230], in their review of methods for software selection, describe three studies that have considered the choice between alternative decision support software applications. Le Blanc and Tawfik Jelassi [231] proposed, within a wider procedure for the evaluation of decision support systems, the use of MCDA to select between applications. The proposed MCDA is based on four types of criteria: technical requirements (e.g. software and hardware compatibility) , functional requirements (e.g. user friendliness), availability of support documentation and training materials, and vendor information (e.g. availability of vendor support for installation and training in the use of the software) [231]. Ossadnik and Lange [232] used an AHP model to choose between three software applications implementing AHP (namely, Automan, *Expert Choice* Pro, and HIPRE+). The criteria used in the software selection software were: 1) criteria of software performance (including e.g. sub-criteria of technical functionality and sub-criteria of functions of rationalisation of usage) and 2) software cost criteria (including the costs of acceptance to users and the initial investment in acquiring the software) [232]. Phillips-Wren et al. [233] proposed a multi-criteria approach for the evaluation of decision support systems based on two types of criteria. The first type of criteria were outcomes of the decision-making process, e.g. proficiency in the phases of decision-making (i.e. intelligence, design, choice and implementation), proficiency in the steps of decision-making (from recognising the decision problem to making a choice) and changes in the organisation or in the decision maker (e.g. reducing the time required for making decisions) [233]. The second type of criteria were outcomes of the decision, e.g. measures of performance in the organisation such as lower costs or increased profit [233]. Dowie et al. [19] have proposed the Decision Resource-Decision Effectiveness Analysis (DRDEA) framework. According to DRDEA [19], the question of choosing between alternative decision technologies (i.e. ways of making decisions), where decision support software applications are particular decision technologies [19], is a multi-criteria decision problem specific to the particular decision situation at hand. The decision problem can be represented by two types of criteria: 1) decision resource criteria expressing the resource requirements associated with using each decision technology (e.g. time required, cognitive effort required, or financial cost [19]) and 2) decision effectiveness criteria expressing the benefits of

using each decision technology (e.g. theoretical grounding, evidential strength and coverage, transparency [19]). In DRDEA, the decision of choosing between alternative decision technologies is preference sensitive and the appropriate analytical technique for solving this decision is MCDA [19].

Le Blanc and Tawfik Jelassi [231], Ossadnik and Lange [232], Phillips-Wren et al. [233] and Dowie et al. [19] all highlight the use of MCDA as an appropriate procedure for choosing between decision support software applications. With respect to the criteria considered relevant for the choice in each of the four studies, the DRDEA framework is the only method that proposes an explicit distinction between two groups of criteria: 1) the benefits (i.e. decision effectiveness) and 2) the resource requirements or costs (i.e. decision resources of using these applications) of using these software applications to make decisions. The application of the DRDEA framework to specific decision contexts and decision processes allows to explore the trade-offs between these benefits and costs.

9. Key issues.

From the above literature review, a number of key issues of relevance to this thesis study have been identified:

1. Although consensus IPDAS guidelines exist regarding the content and the development process of PDAs, these guidelines have been criticised for their limitations. Furthermore, they do not explicitly address the content and development process of MCDA-based PDAs. More specifically, these guidelines do not address the development of MCDA-based PDAs for complex decisions by clinicians within the constraints of their day-to-day clinical practice;

2. Studies have shown that patients are willing and able to use some MCDA methods to make clinical decisions. Other studies have shown that clinicians find some MCDA methods easy to use and potentially useful for helping patients make decisions. In the few existing RCTs evaluating the impact of MCDA-based PDAs, compared to usual consultations these tools can be effective in terms of improving the quality of decision-making. However, different MCDA methods differ greatly in cognitive complexity. Considering 1) that the implementation of PDAs in routine clinical practice has not been achieved to the desired extent and 2) that barriers to PDA implementation include concerns about the adequacy of PDAs for some patient

groups (e.g. those with low health literacy) and competing demands and time constraints by clinicians, the use of differing MCDA methods in the development and implementation of MCDA-based PDAs for complex health care decisions in routine clinical practice should be explored;

3. There are a large number of alternative software applications (i.e. templates) that may be used for developing and implementing MCDA-based PDAs. These applications can be more or less fit-for-purpose depending on the specific decision context and decision process at hand. According to the Decision Resource-Decision Effectiveness Analysis (DRDEA), the choice between alternative templates may be expressed in terms of trade-offs between decision resource (DR) and decision effectiveness (DE) criteria using MCDA. This framework has not been tested in the context of the development and implementation by clinicians of MCDA-based PDAs for complex decisions in routine clinical practice.

Based on the above key issues, the aim and objectives of this thesis are described below.

Aim: To analyse, as proof of concept, the use of MCDA for the development, implementation and evaluation of interactive patient decision aids in routine clinical practice

Objectives:

1. To assess the use with clinicians in the Spanish NHS of two alternative MCDA software applications which implement dissimilar MCDA techniques in the development of a PDA in routine clinical practice;

2. To assess the use with clinicians in the Spanish NHS of the same two alternative MCDA software applications in the implementation of a PDA in an environment replicating actual clinical consultations;

3. To build a meta-multi-criteria decision model based on the DRDEA framework and assess the use of this model by clinicians in the Spanish NHS to make the choice between the two MCDA applications as the basis for a PDA.

10. The case: a patient with lung cancer.

This proof of concept study is built around using MCDA in the development, implementation and evaluation of interactive patient decision aids in routine clinical practice. To put the study into practice with clinicians in the Spanish NHS, a case in the general field of pulmonology was developed to be the subject of the MCDA-based PDA. To establish the case, an interview was held with an experienced pulmonologist from a hospital (hospital H1) in the Spanish NHS. The main question posed to the pulmonologist was to identify a hypothetical patient who would face a complex decision which was of relevance in routine clinical practice in Spanish hospitals. The clinician identified the following hypothetical patient:

- a 69 year-old patient recently diagnosed with stage $IIIA_3$ non-small cell lung cancer (TNM stage T2N2M0 [234]) and with the following comorbidities: 1) light chronic obstructive pulmonary disease (COPD) with a good lung function; 2) myocardial infarction three years back treated with stent surgery.

This case is complex because the treatment is not clear. According to the clinical guidelines for the treatment of non-small cell lung cancer patients developed in hospital H1 [235], and which the pulmonologist shared with this researcher, the recommended treatment for a patient with Stage $IIIA_3$ T2N2M0 non-small cell lung cancer depends on the patient's age and lung function. If the patient is under 70 and has a good lung function, the recommended active treatment is neoadjuvant chemotherapy with resective intent. However, if the patient is over 70 and with a poor lung function, the recommended active treatment is concurrent chemo-radiotherapy. The patient in this case study inhabits a blurred area between both recommendations. The decision is then preference-sensitive and the use of a PDA to make this decision is appropriate.

Chapter 2: Using *Expert Choice* and *Annalisa in Elicia* to develop, with clinicians in the Spanish NHS, a patient decision aid for the clinical management of Stage IIIA3 non-small cell lung cancer.

This chapter explains the methods and results related to Objective 1 of this thesis, i.e.:

- To assess the use with clinicians in the Spanish NHS of two alternative MCDA software applications which implement dissimilar MCDA techniques in the development of a PDA in routine clinical practice.

The chapter begins with a justification for the choice of software applications (from now on referred to simply as "templates") *Expert Choice* and *Annalisa in Elicia* (*ALEL*) for the development of a Stage $IIIA_3$ non-small cell lung cancer clinical management PDA. It then presents an overview of the methods used to develop the PDA using *Expert Choice* in hospital H1 and *ALEL* in hospital H2. The chapter then describes step-by-step the development of the two template-based PDAs. The results and the corresponding commentary are presented at the end of the chapter.

1. Rationale for choosing *Expert Choice* and *ALEL* as the basis for the lung cancer management PDA.

The author of this research chose, for the development and implementation of the lung cancer management patient decision aid (PDA), two templates. The selection was made on the basis of three criteria. First, the two templates should implement methods previously explored in medical decision-making situations. Second, the two templates should be relatively easy to use. Third, with a view towards subsequent evaluation using Decision Resources-Decision Effectiveness Analysis (DRDEA), the two templates should differ in terms of the likely benefits (decision effectiveness) and the likely costs (decision resources).

1.1. *Expert Choice.*

The first chosen template is *Expert Choice*. *Expert Choice* implements the Analytic Hierarchy Process (AHP) MCDA approach. The use of the AHP in medical decision making has been described in several studies, e.g. in the choice of best screening test

for colorectal cancer [18, 179], the choice of antimicrobial therapy for pyelonephritis [236], and the choice of whether or not to undergo prostate cancer screening [194]. The AHP has been described as useful for patient decision support due to the ease of use of its pair-wise comparisons elicitation procedure and due to the strength of measurement of the AHP methodology [116].

1.2. *ALEL.*

The second chosen template is *ALEL*. *ALEL* implements SAW, a simple method, via the software Annalisa. The use of SAW via Annalisa in medical decision making has been described e.g. in Masya et al. [175] for the choice of colorectal cancer treatment and by Cunich et al. [176] for deciding whether or not to take a prostate screening test. *ALEL* was explicitly designed to be of practical use [19].

1.3. Different benefits and costs in *Expert Choice* and *ALEL*.

Expert Choice and *ALEL* differ between them in the likely benefits (i.e. decision effectiveness) and costs (i.e. decision resources) associated with their use in the development and delivery of a PDA. First of all, *Expert Choice* is a hierarchical MCDA template, while *ALEL* is not. In this sense, under the assumption that for a complex clinical decision question such as the choice of lung cancer treatment the decision model is highly likely to be a hierarchy of decision criteria, *Expert Choice* has the benefit over *ALEL* that it allows for a more thorough representation of all the aspects of the decision. Other things being equal, this is likely to come at a cost. A hierarchical PDA is likely to take more time to deliver in a clinical consultation than a non-hierarchical PDA. Second of all, the procedure used to elicit patient preferences with *Expert Choice* (pair-wise comparisons of relative importance of the criteria) has higher costs and higher benefits than that used with *ALEL* (direct weight elicitation). On the cost side, the pair-wise comparisons with *Expert Choice* are likely to require more time than the direct criteria weight elicitation using *ALEL*. On the benefit side, the AHP verifies the consistency of the user's pair-wise comparison judgments, which is not possible with direct weight elicitation in *ALEL*.

In summary, *Expert Choice* and *ALEL* have a track record in medical decision-making, both are easy to use and both differ between them in the likely benefits (decision effectiveness) and costs (decision resources). The next section explains the

methods used to develop the lung cancer management PDA using *Expert Choice* and using *ALEL*.

2. Overview of methods.

This section describes how each of the two MCDA templates (*Expert Choice* and *ALEL*) was used to develop a lung cancer clinical management PDA with a team of clinicians from a hospital in the Spanish NHS. The two chosen hospitals were:

- Hospital Germans Trias I Pujol (from now on hospital H1), in Badalona, a large town in the outskirts of Barcelona

- Hospital Reina Sofia (from now on hospital H2), in Cordoba.

In each of the two hospitals, three clinicians were recruited by the present author: one pulmonologist (referred throughout as clinician 1), one oncologist (referred throughout as clinician 2) and one thoracic surgeon (referred throughout as clinician 3). The rationale for this choice is that pulmonologists, thoracic surgeons and oncologists are the three medical specialists who are typically most heavily involved in the clinical management of non-small cell lung cancer patients. It was deemed important to incorporate the clinical perspectives of these three clinical specialisations in the process of PDA development.

The *Expert Choice*-based PDA was developed in close collaboration with the team of three clinicians recruited from hospital H1. The *ALEL*-based PDA was developed in close collaboration with the team of three clinicians recruited from hospital H2.

Recall from Chapter 1 that the hypothetical patient is a 69 year-old patient recently diagnosed with stage $IIIA_3$ non-small cell lung cancer (TNM stage T2N2M0) and with the following comorbidities: 1) light chronic obstructive pulmonary disease (COPD) with a good lung function; 2) myocardial infarction treated three years back with stent surgery.

The development of the *Expert Choice*-based PDA in hospital H1 and the *ALEL*-based PDA in hospital H2 was divided into the following three stages of the MCDA process:

STAGE 1. Determining the relevant options for the clinical management of the hypothetical lung cancer patient;

STAGE 2: Determining the criteria relevant to the patient for choosing between the options;

STAGE 3: Calculating the scores of the options on the criteria;

The processes associated with developing the *Expert Choice*-based PDA (in H1) and the *ALEL*-based PDA (in H2) were strongly context led. Aiming at generating knowledge in the practical context of day-to-day clinical practice, these processes were led by the present author, but evolved based on the opinions of the clinicians and in response to the constraints confronted by these clinicians. One constraint that shaped these processes was the limited time that the clinicians had to participate in this research project due to their highly demanding clinical work.

3. Developing the *Expert Choice*-based PDA and the *ALEL*-based PDA.

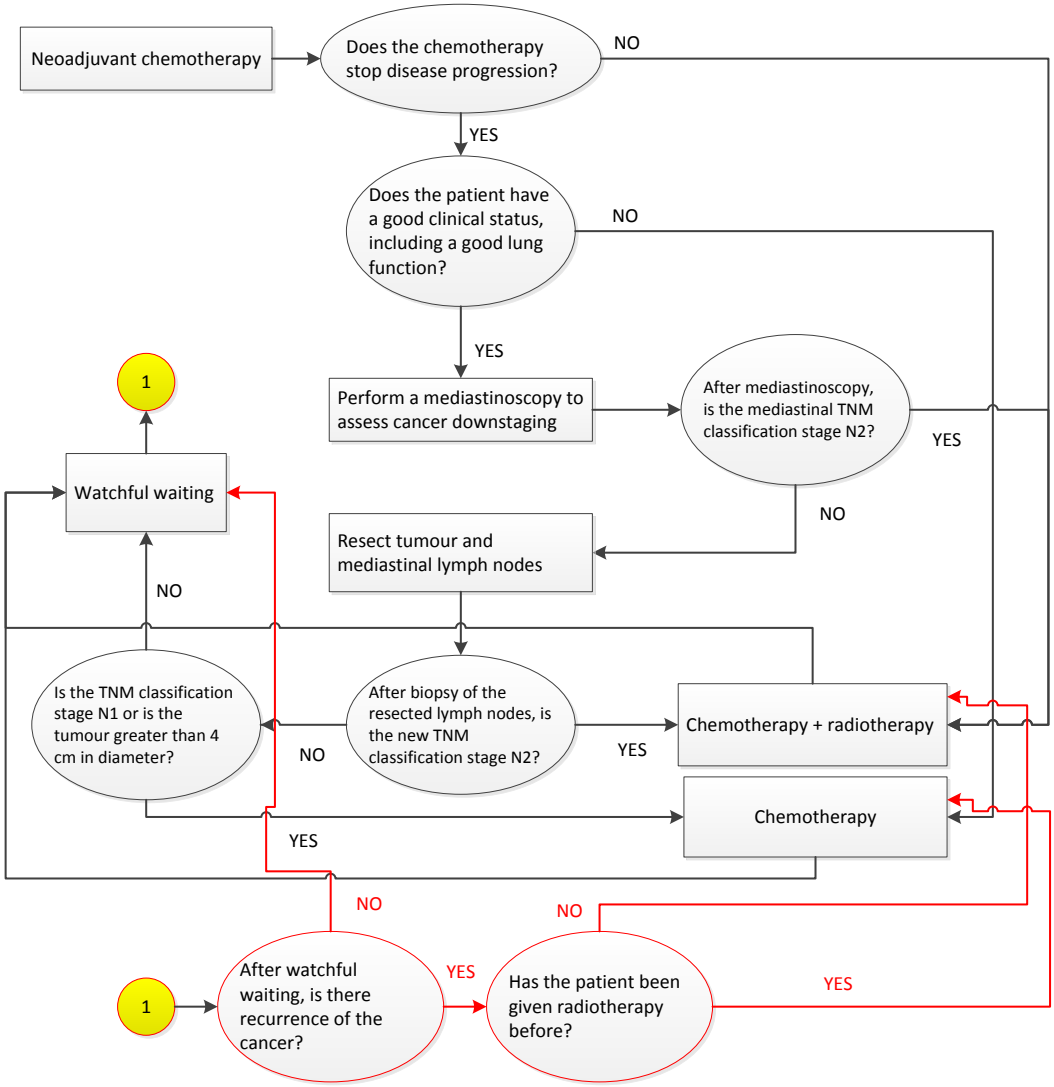
3.1. STAGE 1: determining the options for the clinical management of the hypothetical lung cancer patient

This stage was undertaken by the present author, based on 1) a review of non-small cell lung cancer clinical guidelines [235, 237] and 2) on extensive one-on-one discussions with the pulmonologists from hospitals H1 and H2. The options are common to both the *Expert Choice*-based PDA (in H1) and to the *ALEL*-based PDA (in H2). They are:

Option A_1: neoadjuvant chemotherapy with resective intent.

With this option, the hypothetical patient is initially treated with neoadjuvant chemotherapy with the aim that the NSCLC is down-staged so that the patient's tumour can be surgically removed. The chemotherapy is termed "neoadjuvant" because it is given prior to the main treatment. The main treatment is tumour resection in the form of lobectomy if a number of conditions are met after the patient has been given the neoadjuvant chemotherapy. Figure 2.1 illustrates the pathway of possible clinical interventions associated with this option. The elements of the pathway highlighted in red indicate cancer recurrence.

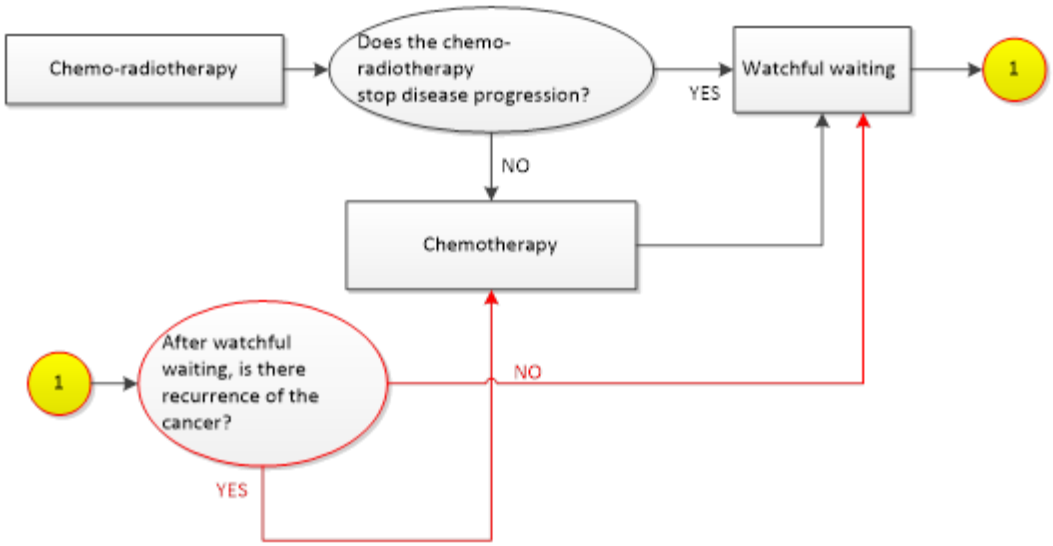
Figure 2.1. Pathway of possible clinical interventions under option A_1, neoadjuvant chemotherapy with resective intent.



Option A_2: concurrent chemo-radiotherapy

With this option, the hypothetical patient is not considered for surgery. He is initially treated with concurrent-chemotherapy with the aim to stop disease progression. The subsequent interventions under this option are dependent on the uncertain success of the chemo-radiotherapy at reversing the cancer. Figure 2.2 illustrates the pathway of possible clinical interventions associated with this option. The elements of the pathway highlighted in red indicate cancer recurrence.

Figure 2.2. Pathway of possible clinical interventions under option A_2, concurrent chemo-radiotherapy.



Option A_3: best supportive care.

Best supportive care, or palliative care, consists in providing the hypothetical patient with 1) good communication to facilitate decision-making, 2) symptoms control, psychosocial support (during the disease and in the last days of life), and physical care in the last days of life [238].

Apart from options A_1, A_2, and A_3, no additional clinical management options were considered available for the patient by the clinicians in either hospital.

Once the options were determined, the second stage in PDA development was to determine the criteria relevant for the decision for 1) the *Expert Choice*-based PDA and 2) the *ALEL*-based PDA. This stage is described below.

3.2. STAGE 2: Determining the relevant criteria for choosing between the options.

This stage was initially undertaken separately for the *Expert Choice*-based PDA (in H1) and the *ALEL*-based PDA (in H2) during several group meetings between the author of this research and each team of clinicians.

3.2.1. Initial sets of criteria.

In each hospital, the same procedure was initially used to define the criteria for the PDA: the present author met with the three specialists and asked them to agree on a set of criteria that the hypothetical lung cancer patient would consider relevant for the decision. In each hospital the clinicians considered the same five relevant criteria:

1. The duration of life, i.e. the life expectancy;

2. The burden of treatment, i.e. the treatment-related adverse effects;

3. The quality of life in the medium term, where medium term was defined as two years after the start of treatment;

4. The financial burden in the medium term, i.e. the financial problems derived, two years after the start of treatment, from 1) direct expenditures related with the disease and/or the treatment, and from 2) the opportunity cost of not being able to earn a living as a result of being ill;

5. The quality of the health care experience (i.e. those aspects of the health care delivery which are positive for one's well-being as a patient) from the start of treatment until the medium term (i.e. 2 years after the start of treatment).

In each hospital, the team of clinicians defined a set of sub-criteria for: quality of life in the medium term (criterion 3 above), and quality of the health care experience (criterion 5 above). For quality of life in the medium term, in both hospitals the team of clinicians opted for using the items of the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 questionnaire [239, 240] as the basis for the relevant sub-criteria. For quality of the health care experience, the team of clinicians using *Expert Choice* opted to define the following four sub-criteria: 1) visits to the health services/ hospital inpatient stays, 2) waiting time due to waiting lists between interventions, 3) duration of treatment by the same team of clinicians, 4) personalised care. For quality of the health care experience, the team of clinicians using *ALEL* opted, given the impossibility of using a hierarchy of criteria in *ALEL*, to represent this dimension using only one criterion. The chosen criterion was: visits to

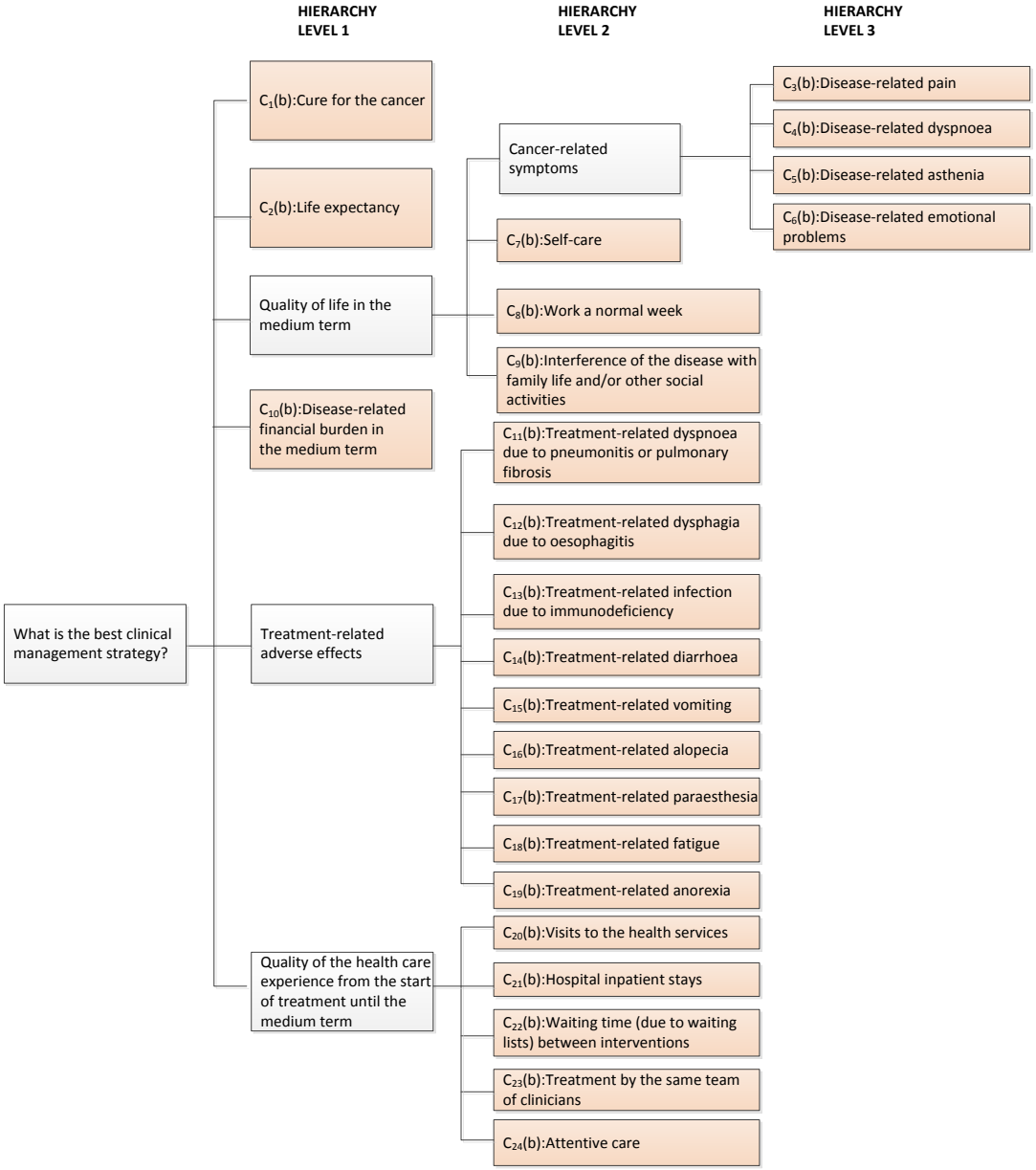
the health services/ hospital inpatient stays from the start of treatment until the medium term (i.e. two years after starting the treatment).

The criteria hierarchies initially built with each team of clinicians (shown in Appendix 1) relied greatly in the items of the EORTC QLQ-C30 questionnaire. The present author considered this was problematic. First, the EORTC QLQ-C30 instrument uses interval scales to score its items. These scores are incompatible with the AHP which requires scores to be measured on ratio scales. It was considered confusing to provide ratio scores for the same items which are measured on interval scales in an instrument which is furthermore subject to copyright. Second, the author of this researcher considered that many of the criteria of these initial hierarchies were not mutually preference independent. After discussing this issue in a group meeting with each team, in both cases the initial hierarchy was discarded.

3.2.2. Final set of criteria.

Although in each hospital the initial hierarchy of criteria proposed by each team clinicians was discarded, the information contained in each hierarchy was used by the present author to build a single hierarchy for both the *Expert Choice*-based PDA (in H1) and the *ALEL*-based PDA (in H2). This hierarchy was presented to each of the two teams of clinicians in a further group meeting and unanimously considered an adequate representation of the decision problem. The hierarchy is presented in Figure 2.3 (see Appendix 2 for a definition of the items in the hierarchy).

Figure 2.3. Final hierarchy of criteria for the Expert Choice-based PDA and the ALEL-based PDA



In Figure 2.3, note that the goal is choosing the best clinical management strategy for the hypothetical patient. The criteria shaded in grey are top-level criteria composed of sub-criteria. The criteria shaded in light red are bottom-level criteria of the hierarchy. There are twenty-four bottom-level criteria, each with the mathematical notation $C_k(b)$ $(k = 1,2 \dots 24)$.

3.2.3. Representing the hierarchy of criteria in the PDA.

At this stage of the development of the two PDAs, the issue arose of representing the hierarchy of criteria during the delivery of the PDA in consultations. Both teams of clinicians considered that the hierarchy was too large to display fully to a patient. In both teams of clinicians it was agreed that only the top-level criteria (i.e. the criteria on Level 1 of the hierarchy) would be represented in the PDA.

Once the criteria were defined and the clinicians agreed to represent in the PDA only the top-level criteria of the hierarchy, the final stage in PDA development was to calculate the scores of the three options on the top-level criteria for 1) the *Expert Choice*-based PDA in hospital H1 and for 2) the *ALEL*-based PDA in hospital H2. This stage is described below.

3.3. STAGE 3: Calculating the scores of the options on the top-level criteria.

This stage was undertaken separately for the *Expert Choice*-based PDA (in H1) and for the *ALEL*-based PDA (in H2). They are presented separately below.

3.3.1. STAGE 3 for the *Expert Choice*-based PDA: Calculating the scores (or priorities) of the options on the top-level criteria of the hierarchy.

Recall from Chapter 1 that to avoid rank reversals, the ideal AHP should be used whenever options can be added to or removed from the option set. In this research project, the ideal AHP is used to allow for the eventuality of adding or removing options.

Two steps were required using the ideal AHP to calculate the scores (from now on termed priorities for consistency with AHP nomenclature) of the three clinical management options on the top-level criteria of the hierarchy. The first step was calculating the priorities of the options on the bottom-level criteria of the hierarchy. The second step was assigning weights to these bottom-level criteria so that the bottom-criteria priorities can be propagated up the hierarchy. These two steps are described below.

3.3.1.1. Calculating the priorities of the options on the bottom-level criteria for the *Expert Choice*-based PDA.

The calculation of the priorities of the options on the bottom-level criteria required two additional steps: 1) deciding with the three clinicians in hospital H1 how to measure the priorities of the options on the bottom-level criteria, and 2) actually calculating these priorities. The first step is presented in section 3.3.1.1.1. The second step is presented in section 3.3.1.1.2.

3.3.1.1.1. Deciding how to calculate the priorities of the options on the bottom-level criteria for the *Expert Choice*-based PDA.

In *Expert Choice* using the ideal AHP, the calculation of the priorities of the options on a particular bottom-level criterion can be done using two approaches: 1) making pair-wise comparisons of judgments of relative performance between options on that criterion, or 2) directly entering data in the appropriate format. In the first case, the priorities of the options on each criterion are calculated using the eigenvector method. In the second case, the priorities are obtained in two steps: 1) measuring the performance levels of the options on the criteria, 2) transforming these levels into 0-1 priorities on a ratio scale. In both cases, each priority of each option on each criterion is then idealised - that is, normalised by the priorities of the highest performing option on that criterion.

It was proposed by the present author and agreed in a group meeting with the three clinicians in hospital H1 to calculate the priorities of the three options on the twenty-four bottom-level criteria of the hierarchy (i.e. the criteria shaded in light red in Figure 2.3) using both approaches described in the previous paragraph. Pairwise comparisons of relative performance between options were to be used with four criteria: $C_3(b)$-$C_6(b)$. Direct data were to be estimated for the remaining twenty criteria: $C_1(b)$, $C_2(b)$, and $C_7(b)$-$C_{24}(b)$). Table 2.1 describes the mechanics of the pair-wise comparisons for the first set of criteria; Table 2.2 describes the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on the second set of criteria, i.e. the criteria $C_k(b)$ $(k = 1,2,7,8 \ldots 24)$. In Table 2.2, note that criteria $C_{11}(b)$-$C_{19}(b)$ are all criteria reflecting what the clinicians in hospital H1 considered were the main adverse effects of the options. To define the possible levels

of these criteria, it was agreed to use the grading system of the National Cancer Institute Common Terminology Criteria for Adverse Effects (CTCAE) [241] in its version 4 [242]. Table 2.1 and Table 2.2 are self-explanatory and will not be explained further in the main body of the text.

Table 2.1. Description of the pair-wise comparisons used to obtain the priorities of the options on criteria $C_3(b)$-$C_6(b)$

<table>
<tr><th>Criterion $C_k(b)$</th><th>Description of the pair-wise comparisons of relative dominance of the options A_i $(i = 1 \ldots 3)$ with respect to each criterion</th></tr>
<tr><td>Disease-related pain
$C_3(b)$</td><td rowspan="4">The matrix of comparative consequence judgments for each criterion $C_k(b)$ is:
$$CM(C_k(b)) = \begin{pmatrix} c(k)_{A_1,A_1} & c(k)_{A_1,A_2} & c(k)_{A_1,A_3} \\ c(k)_{A_2,A_1} & c(k)_{A_2,A_2} & c(k)_{A_2,A_3} \\ c(k)_{A_3,A_1} & c(k)_{A_3,A_2} & c(k)_{A_3,A_3} \end{pmatrix}$$
As was shown in Chapter 1, the elements of matrix $CM(C_k(b))$ are the comparative judgments of dominance $c(k)_{A_i,A_j}$ of every alternative A_i $(i = 1 \ldots 3)$ with respect to every other alternative A_j $(i = 1 \ldots 3)$. The elements of the diagonal of the matrix are all equal to 1. The elements below the diagonal of the matrix are the reciprocals of their corresponding elements above the diagonal. It is necessary to calculate the three following elements above the diagonal of the matrix:
1. $c(k)_{A_1,A_2}$: The comparative judgment of dominance, on a scale between 1 and 9, of option A_1 (neoadjuvant chemotherapy with resective intent) compared to option A_2 (concurrent chemo-radiotherapy) with respect to criterion $C_k(b)$
2. $c(k)_{A_1,A_3}$: The comparative judgment of dominance, on a scale between 1 and 9, of option A_1 (neoadjuvant chemotherapy with resective intent) compared to option A_3(best supportive care) with respect to criterion $C_k(b)$
3. $c(k)_{A_2,A_3}$: The comparative judgment of dominance, on a scale between 1 and 9, of option A_2 (concurrent chemo-radiotherapy) compared to option A_3 (best supportive care) with respect to criterion $C_k(b)$</td></tr>
<tr><td>Disease-related dyspnoea
$C_4(b)$</td></tr>
<tr><td>Disease-related asthenia
$C_5(b)$</td></tr>
<tr><td>Disease-related emotional problems
$C_6(b)$</td></tr>
</table>

Table 2.2. Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion**
Cure $C_1(b)$	Probability that there is no tumour activity 5 years after starting treatment	Min: 0 Max: 1	$x_{i,1} \in [0,1] \subset \mathbb{R}^+$
Life Expectancy $C_2(b)$	Survival in years	Min: 0 years Max: $LE(max)$, which is the highest estimated survival under the most efficacious of the three options	$x_{i,2} \in [0, LE(\max)] \subset \mathbb{R}^+$
Self-care $C_7(b)$	Probability that the patient will, in the medium term, be able to take care of himself without help from others	Min: 0 Max: 1	$x_{i,7} \in [0,1] \subset \mathbb{R}^+$
Work a normal week $C_8(b)$	Probability that the patient will, in the medium term*, be able to work a normal week (i.e. 40 hours in a week)	Min: 0 Max: 1	$x_{i,8} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion**
Interference of the disease with family life or with social activities $C_9(b)$	Probability that the disease will, in the medium term*, interfere from moderately to extremely in the patient's family life and in his family and social relations	Min: 0 Max: 1	$x_{i,9} \in [0,1] \subset \mathbb{R}^+$
Disease-related financial burden in the medium term* $C_{10}(b)$	Probability that the disease will, in the medium term*, cause the patient form moderate to severe financial difficulties	Min: 0 Max: 1	$x_{i,10} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related dyspnoea as a consequence of pneumonitis or pulmonary fibrosis $C_{11}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of dyspnoea as a consequence of treatment-related pneumonitis and/or pulmonary fibrosis: - grade 2 (dyspnoea with minimal exertion) - grade 3 (dyspnoea at rest) - grade 4 (dyspnoea with life-threatening consequences)	Min: 0 Max: 1	$x_{i,11} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related dysphagia as a consequence of oesophagitis $C_{12}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of dysphagia as a consequence of treatment-related oesophagitis: - grade 2 (symptomatic and altered eating/swallowing) - grade 3 (severely altered eating/swallowing requiring tube feeding) - grade 4 (dysphagia with life-threatening consequences)	Min: 0 Max: 1	$x_{i,12} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related infection due to immunodeficiency $C_{13}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of treatment-related infection as a consequence of immunodeficiency: - grade 2 (requiring antibiotics) - grade 3 (requiring, in addition to antibiotics, radiologic or operative intervention) - grade 4 (infection with life-threatening consequences)	Min: 0 Max: 1	$x_{i,13} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each conseq.**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related diarrhoea $C_{14}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of treatment-related diarrhoea: - grade 2 (increase of 4 to 6 stools per day over baseline) - grade 3 (increase of 7 or more stools per day over baseline, incontinence, limiting self-care, requires hospitalisation) - grade 4 (diarrhoea with life-threatening consequences)	Min: 0 Max: 1	$x_{i,14} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each conseq.**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related vomiting $C_{15}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of treatment-related vomiting: - grade 2 (3-5 episodes separated by 5 minutes in 24 hours) - grade 3 (6 or more episodes separated by 5 minutes in 24 hours, requires tube feeding and hospitalisation) - grade 4 (vomiting with life-threatening consequences)	Min: 0 Max: 1	$x_{i,15} \in [0,1] \subset \mathbb{R}^+$
Treatment-related alopecia $C_{16}(b)$	Probability that the patient will have the following moderate grade of treatment-related alopecia*: - grade 2 (loss of 50% or more of one's hair associated with psychosocial impact)	Min: 0 Max: 1	$x_{i,16} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each conseq.**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related paraesthesia $C_{17}(b)$	Probability that the patient will have any of the following moderate to severe grades of treatment-related paraesthesia*: - Grade 2 (moderate paraesthesia) - Grade 3 (severe paraesthesia limiting self-care)	Min: 0 Max: 1	$x_{i,17} \in [0,1] \subset \mathbb{R}^+$
Treatment-related fatigue $C_{18}(b)$	Probability that the patient will have any of the following moderate to severe grades of treatment-related fatigue*: - Grade 2 (fatigue not relieved by rest) - Grade 3 (fatigue not relieved by rest, limiting self-care)	Min: 0 Max: 1	$x_{i,18} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each conseq.**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Treatment-related anorexia $C_{19}(b)$	Probability that the patient will have any of the following moderate to extremely severe grades of treatment-related anorexia: - Grade 2 (oral intake altered without significant weight loss or malnutrition requiring oral nutritional supplements) - Grade 3 (significant weight loss or malnutrition, requiring tube feeding) - Grade 4 (anorexia with life threatening consequences	Min: 0 Max: 1	$x_{i,19} \in [0,1] \subset \mathbb{R}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each conseq.**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on each criterion.**
Visits to the health services $C_{20}(b)$	Total number of cancer-related visits to any outpatient health service from the start of treatment until the medium term**	Min: 0 Max: $V(max)$, which is the highest number of visits incurred in with any of the three options	$x_{i,20}$ $\in \{1,2, \dots V(\max)\}$ $\subset \mathbb{N}^+$
Hospital in-patient stays $C_{21}(b)$	Total number of days spent by the patient in the hospital due to cancer-related hospitalisation from the start of treatment until the medium term**	Min: 0 Max: $D(max)$, which is the highest number of days spent in hospital due to a cancer-related hospitalisation out of any of the three options	$x_{i,21}$ $\in \{1,2, \dots D(\max)\}$ $\subset \mathbb{N}^+$

* In the CTCAE version 4 system severity grades higher than grade 3 are not defined for fatigue

** The medium term is defined as 2 years after the start of treatment

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on the criterion.**
Waiting time (due to waiting lists) between interventions $C_{22}(b)$	Total number of days over the optimal calendar required to continue treating the patient due to waiting lists from the start of treatment until the medium term*	Min: 0 Max: $W(max)$, which is the highest number of days required to continue treating the patient due to waiting lists from the start of treatment until the medium term out of any of the three options	$x_{i,22} \in \{1,2, \dots W(\max)\}\mathbb{N}^+$

Table 2.2 (cont.). Description of the variables used to measure the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on the criterion.**
Treatment by the same team of clinicians $C_{23}(b)$	Whether or not the patient will be treated by the same clinician or by the same team of clinicians from the start of treatment until the medium term*: No=0 Yes=1	Min: 0 (No) Max: 1 (Yes)	$x_{i,23} \in \{0,1\} \subset \mathbb{N}$
Attentive care $C_{24}(b)$	Whether or not the patient will always be treated by his clinician(s) in a caring and considerate fashion from the start of treatment until the medium term: No=0 Yes=1	Min: 0 (No) Max: 1 (Yes)	$x_{i,24} \in \{0,1\} \subset \mathbb{N}$

* The medium term is defined as 2 years after the start of treatment.

Figure 2.4 shows the transformation of the pairwise comparisons of relative performance between the options into priorities using the eigenvector method. Figure 2.5 shows the transformations of the performance levels $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$ into priorities $s_{i,k}$. The transformations from Figure 2.4 are self-explanatory and will not be explained further in the main body of the text. The transformations from Figure 2.5 are explained immediately after Figure 2.5 in the main body of the text.

Figure 2.4. Calculation of the priorities $s_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 3,4,5,6)$ (Expert Choice-based PDA)

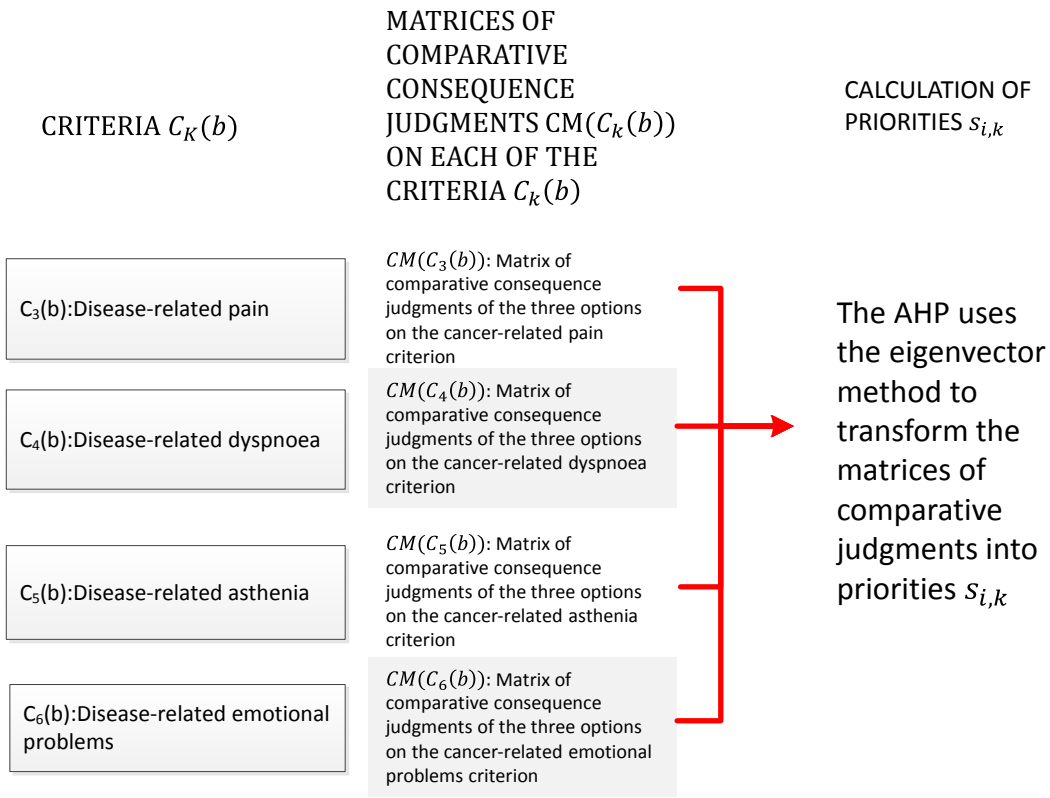


Figure 2.5. Transformation of the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$ into priorities $s_{i,k}$ (Expert Choice-based PDA)

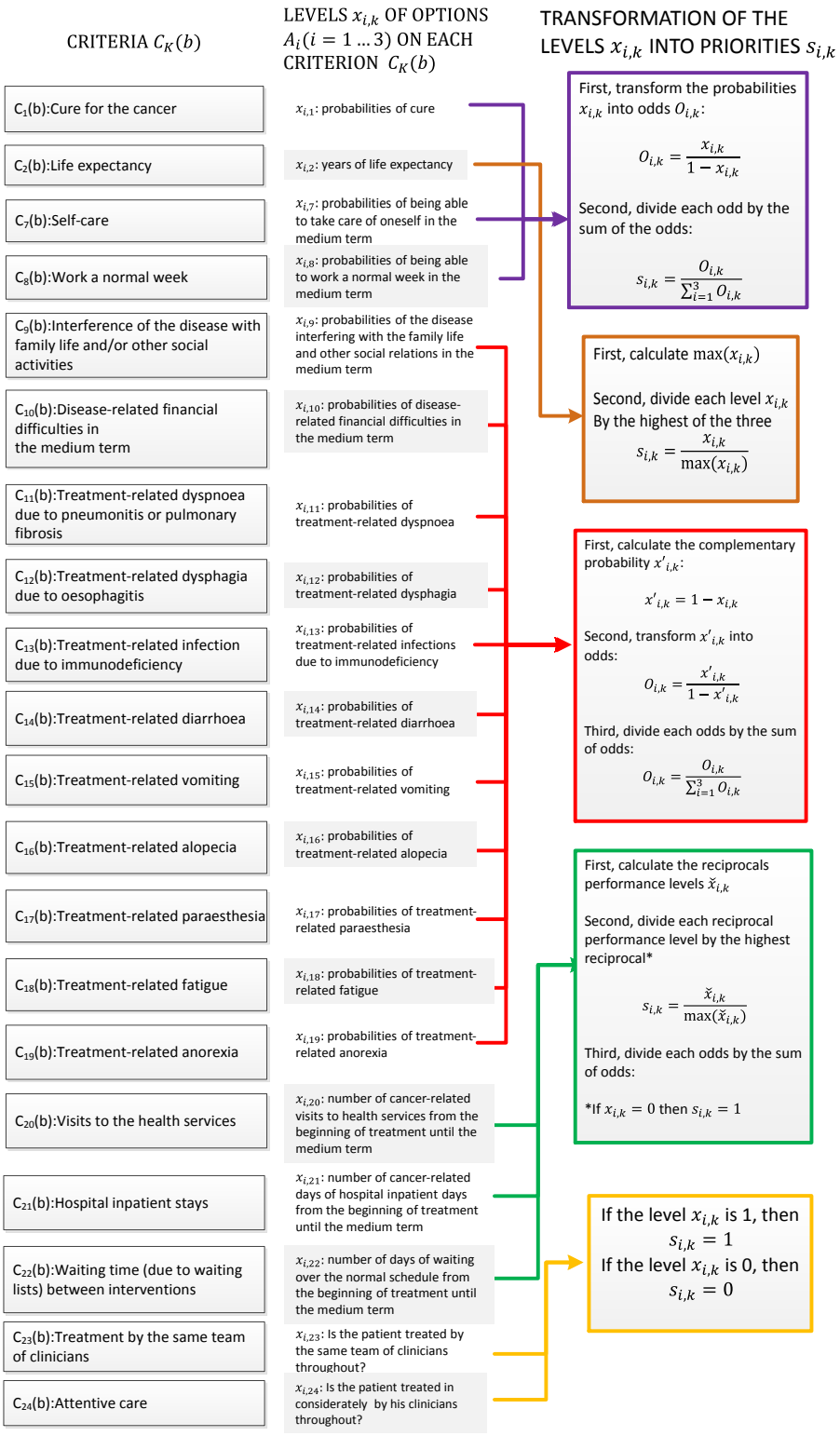


Figure 2.5 describes the calculations for transforming the performance levels $x_{i,k}$ of the three clinical management options A_i $(i = 1,2,3)$ on the criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$ into priorities $s_{i,k}$ on a ratio scale, which was the type of scale required by the AHP. From Figure 2.5 note five types of transformations, each highlighted with a colour code. Each of these transformations is explained below:

1. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 1,7,8)$, respectively cure for cancer, being able to take care of oneself in the medium term (i.e. two years after the start of treatment), and being able to work a normal week in the medium term (i.e. two years after the start of treatment), colour coded in purple in Figure 2.5. These performance levels were, respectively, the probability of cure, the probability of being able to take care of oneself in the medium term, and the probability of being able to work a normal week in the medium term under each clinical management strategy. These probabilities needed to be transformed into priorities on a ratio scale so that a higher priority indicated a better outcome and a lower priority indicated a worse outcome. To transform probabilities into priorities closely resembling the AHP ratio scale, the odds related with these probabilities were then calculated, as indicated by Dolan [236]. With the above odds calculations, the transformation of 1) a probability of avoiding adverse effects equal to 0 and of 2) a probability of avoiding an adverse effect equal to 1 into odds was not possible. To circumvent this issue, it was decided to assign probabilities of 0 a very low but positive odds (i.e. 0.001) and to assign probabilities of 1 an odds close to 1 but below 1 (0.999). The procedure is explained in the purple box in Figure 2.5;

2. Transformation of performance levels $x_{i,k}$ on criterion $C_2(b)$, life expectancy, colour-coded in brown in Figure 2.5. Life expectancy was measured in years. The transformation of years of life expectancy into priorities on a ratio scale such that a higher priority reflected higher life expectancy and a lower priority reflected lower life expectancy was achieved by dividing the years of life expectancy for each option by the life expectancy of the highest performing option, as shown in the brown box in Figure 2.5;

3. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 9 \dots 19)$, adverse effects of the clinical management options, colour-coded in red in Figure 2.5. These performance levels were all probabilities of adverse effects due to the clinical

management options. They needed to be transformed into priorities on a ratio scale so that a higher priority indicated a better outcome (i.e. less adverse effects) and a lower priority indicated a worse outcome (i.e. more adverse effects). The first step to calculate the priorities for the options was to transform the probabilities of adverse effects the probabilities of avoiding adverse effects. This was achieved by calculating the complementary probabilities, as indicated in the red box in Figure 2.5. The second step involved expressing these probabilities as ratio comparisons by calculating the odds related with these probabilities as was explained in point 1.

4. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 20,21,22)$, visits to the health services, hospital inpatient stays, and waiting time (due to waiting lists) between interventions, colour-coded in green in Figure 2.5. These performance levels were measured respectively as the number of visits to the health services, the number of days in hospital, and the number of days of waiting time (due to waiting lists) between interventions with each option. The transformation of these performance levels into priorities on a ratio scale such that a higher priority reflected a better outcome and a lower priority reflected a worse outcome requires two steps. First, the reciprocals of these performance levels were calculated. Second, they were transformed into priorities by dividing the performance level for each option by the performance of the highest performing option on each criterion, as shown in the green box in Figure 2.5;

5. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 23,24)$, treatment by the same team of clinicians and attentive care (colour-coded in orange in Figure 2.5). These performance levels were measured by dichotomous 1/0 variables (corresponding to statements “yes”/”no”). The transformation of the performance levels of the options into priorities was done simply by assigning a performance level of 1 a priority of 1 and assigning a performance level of 0 a priority of 0, as is shown in the orange box in Figure 2.5.

The above transformations completed the procedure agreed with the clinicians in hospital H1 for scoring the performance of the three clinical management options on the bottom-level criteria for the *Expert Choice*-based PDA. The next step was actually calculating these priorities.

3.3.1.1.2. Calculating the priorities of the options on the bottom-level criteria for the *Expert Choice*-based PDA.

Once the procedure to calculate the priorities of the performance levels of the options – recall, neoadjuvant chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2) and best supportive care (A_3)- on the bottom-level criteria of the hierarchy was agreed with the clinicians, the next step in quantifying the evidence was calculating these priorities. This required two additional steps. First, obtaining estimates for the performance levels $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on each of the bottom-level criteria $C_k(b)$ $(k = 1,2 \dots 24)$. Second, transforming these performance levels into priorities using the procedures outlined in Figure 2.5 and described in the previous section. The first step requires explanation, the second is immediate applying the relevant formulas in Figure 2.5, and will not be further described in the main body of the text.

To calculate the performance levels of the options on the bottom-level criteria, the most evident approach would have been performing a series of literature reviews of the published scientific literature regarding these consequences. This posed one immediate problem. The two teams of clinicians did not have time to undertake any literature reviews due to their busy working schedules. As the research was aimed at generating knowledge regarding the use of MCDA templates for the development and delivery of a PDA in the practical context of day-to-day clinical practice, the present author felt that it was important to take the time constraints of the clinicians into account - hence, it was decided that the methods for developing the *Expert Choice-based* and the *ALEL-based* Stage $IIIA_3$ NSCLC PDAs would not resort to literature reviews as a source of evidence information.

The present author did undertake a literature review of the evidence regarding the levels of the three options on the bottom-level criteria to get a general understanding of the availability of information regarding these consequences. The review resulted in a perhaps unsurprising dearth of evidence (see Appendix 3).

Considering the time constraints that the three clinicians from hospital H1 confronted, the chosen source of the evidence information for the *Expert Choice*-based PDA was the clinicians' judgments, i.e. their expert opinion. These judgments

were elicited for each of the three clinicians - the pulmonologist (clinician 1), the oncologist (clinician 2), and the thoracic surgeon (clinician 3).

To elicit these judgments, a questionnaire were developed by the present author using the *ELICIA* online survey functionality [20].

The questionnaire, termed the *Expert Choice-based PDA evidence generation questionnaire* (EC-EGQ) was composed of 112 questions. It was developed to elicit, from each of the three clinicians in hospital H1, two types of judgments:

- First, judgments about the levels $x_{i,k}$ $(i = 1 \ldots 3; k = 1,2,7 \ldots 24)$ of the options A_i $(i = 1 \ldots 3)$ on the criteria $C_1(b)$, $C_2(b)$, and $C_7(b)$-$C_{24}(b)$

- Second, pair-wise comparisons of the perceived dominance of each of the options A_i $(i = 1 \ldots 3)$ with respect to each other option on criteria $C_3(b)$-$C_6(b)$.

Each of the three clinicians from hospital H1 was asked to fill in the EC-EGQ questionnaire online. Three sets of responses were obtained, one set for each clinician

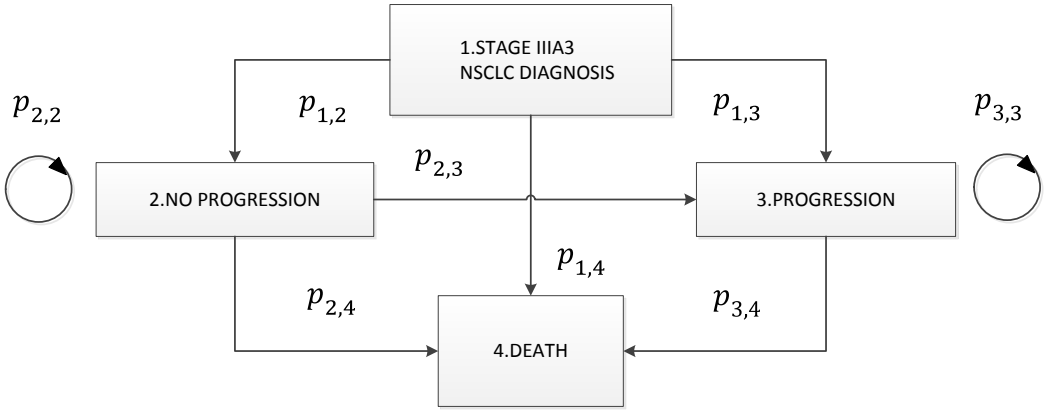
Appendix 4 presents the EC-EGQ and the judgments elicited from each of the three clinicians. For most criteria, the performance levels of the options on the criteria were directly obtained from the questionnaire responses. There was one criterion, however, for which this was not so. In order to measure the levels $x_{i,2}$ $(i = 1 \ldots 3)$ of the options A_i $(i = 1 \ldots 3)$ on criterion life expectancy ($C_2(b)$), the present author decided to use a Markov model. The description of the Markov model and how it was used is provided now.

Markov models are useful tools for modelling disease progression. In a Markov model, a disease (e.g. lung cancer) is represented mathematically by a succession of mutually exclusive health states through which a patient transitions over time until he/she reaches an absorbing state (e.g. death). The basic functioning of a Markov model involves the following steps. First, mapping out the disease states and the possible transitions that a patient can undergo from one state to another; second, calculating the probabilities of these different transitions (i.e. calculating the matrix of transition probabilities); third, based on the matrix of transition probabilities,

simulating a cohort of patients transitioning over time through the different disease states. The average time spent by the cohort of simulated patients transitioning between the different disease states is calculated by the model. This is the average life expectancy of a typical patient with the disease.

A schematic representation of the Markov model used to represent disease progression for a cohort of Stage $IIIA_3$ non-small cell lung cancer patients is shown in Figure 2.6. In the model, each time cycle at which transitions between states occur was set at six months.

Figure 2.6. Schematic representation of the Stage IIIA3 non-small cell lung cancer Markov model



In Figure 2.6, a cohort of patients begins their journey through the model at "Stage $IIIA_3$ NSCLC diagnosis". In the first six months, patients can either transition to a situation of no disease progression (a situation with no changes, with respect to the diagnosis in terms of the extent of the cancerous lesions), transition to a situation of disease progression, or die. If they transition to no disease progression, then every six months after that the patients may either remain in a situation of no disease progression, transition to disease progression, or die. If the patients transition to disease progression, then every six months after that the patients may either remain in a situation of disease progression or die.

The transitions of the cohort of patients between the disease states in Figure 2.6 are governed by the following matrix of transition probabilities TP:

$$TP = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix} = \begin{pmatrix} 0 & p_{1,2} & p_{1,3} & p_{1,4} \\ 0 & p_{2,2} & p_{2,3} & p_{2,4} \\ 0 & 0 & p_{3,3} & p_{3,4} \end{pmatrix}$$

(2.1)

Each element in matrix TP above describes the transitions between the four following states: 1) Stage $IIIA_3$ NSCLC diagnosis, 2) no progression, 3) progression, 4) death.

The matrix of transition probabilities from expression (2.1) is not the same if the patients in the cohort are undergoing neoadjuvant chemotherapy with resective intent (option A_1), concurrent chemo-radiotherapy (option A_2) or best supportive care (option A_3). To calculate the average life expectancy of a patient with Stage $IIIA_3$ non-small cell lung cancer under each option, a matrix of transition probabilities needs to be estimated for each option and then entered into the Markov model.

Estimates of the three matrices of transition probabilities corresponding to the three clinical management options were obtained from each of the three clinicians in hospital H1 using the same online questionnaire (the EC-EGQ) that was used for obtaining the performance levels of the options on the remaining criteria. The questions asked to each of the clinicians, the resulting matrices of transition probabilities and the resulting estimates of life expectancy under each option are shown in Appendix 4.

The use of the Markov model to generate estimates of life expectancy for the hypothetical patient under each clinical management strategy is justified under the assumption that if the matrices of transition probabilities elicited from the clinicians are reasonably accurate, then the life expectancy calculated by the Markov model is a reasonable approximation of the actual life expectancy of the hypothetical patient.

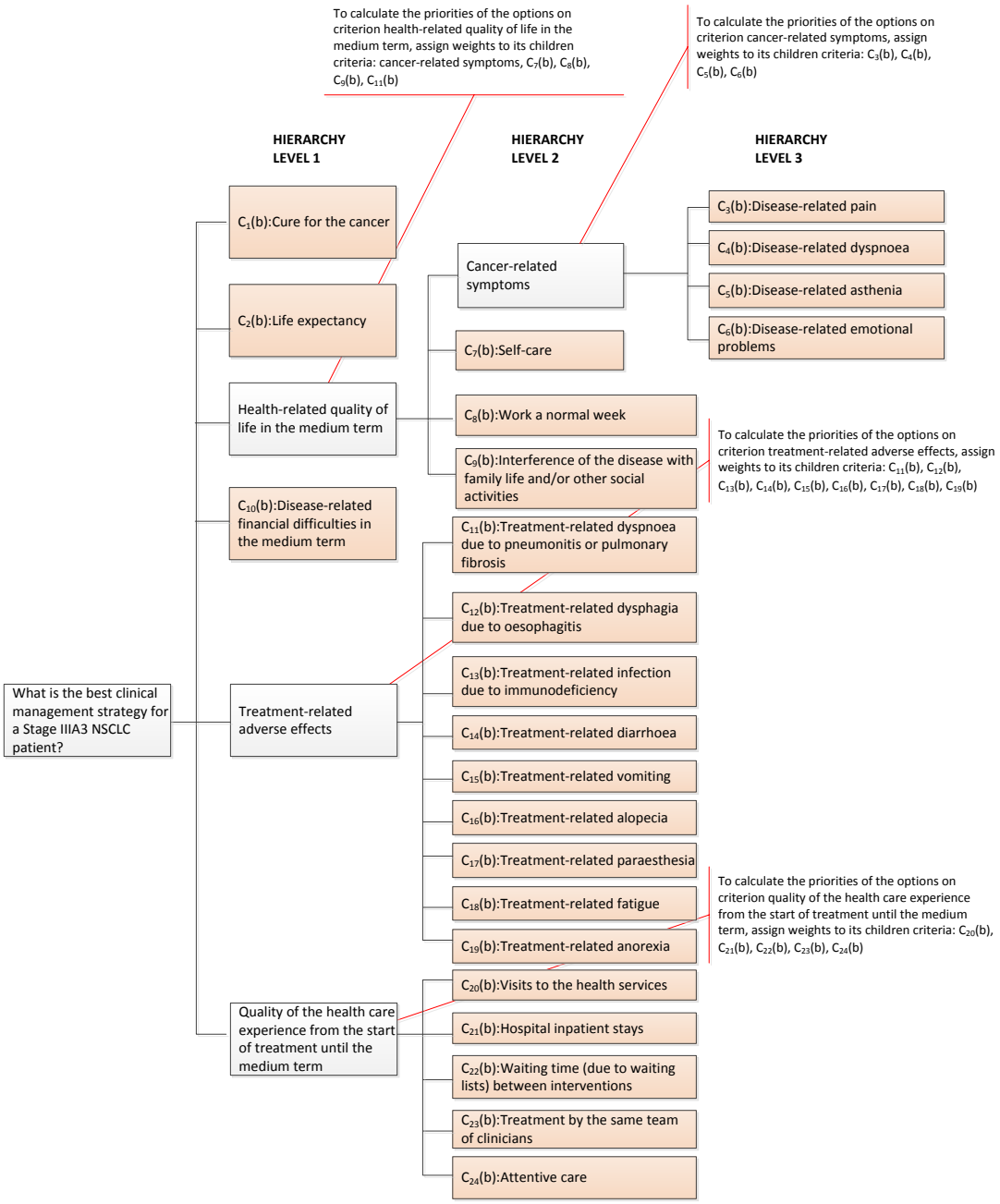
Once all the information on the performance of the three clinical management options on each bottom-level criterion of the hierarchy was elicited from each of the three clinicians in hospital H1, the clinician-specific priorities of the options on each of these criteria were calculated (as specified in section 3.3.1.1.1 and Figure 2.5). These priorities are shown in Appendix 4.

To calculate the priorities of the three options on the top-level criteria of the hierarchy, the next step was propagating the priorities of the options on the bottom-level criteria up the hierarchy. This procedure is explained in the next section

3.3.1.2. Propagating the priorities of the options on the bottom-level criteria up the hierarchy for the *Expert Choice*-based PDA.

Figure 2.7 below shows that three of the criteria in the hierarchy are parent criteria with sub-criteria. They are: cancer-related symptoms (on Level 2 of the hierarchy), quality of life in the medium term (on Level 1 of the hierarchy), treatment-related adverse effects, and quality of the health care experience from the start of treatment until the medium term (Level 1 of the hierarchy). *Expert Choice* automatically calculated the priorities of the options on each parent criterion only once weights have been assigned to all its children criteria. This is illustrated in Figure 2.7.

Figure 2.7. Calculating the priorities of the options on the top-level criteria (Expert Choice-based PDA)

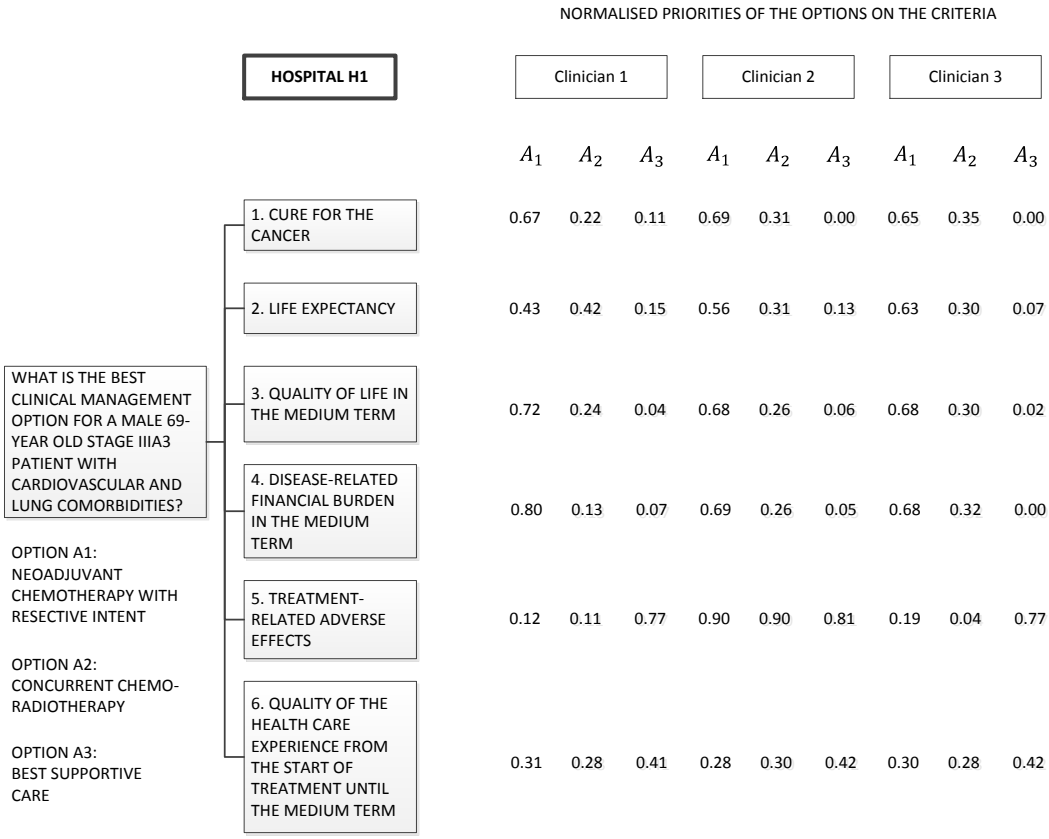


Once the weights were assigned to all the children criteria of every parent criterion, *Expert Choice* calculated the priorities of the options in the top-level criteria by propagating the priorities of the options on these criteria up the hierarchy.

To avoid injecting preferences into the calculations, equal weights were assigned to all the children criteria of each parent criterion. This provided the priorities shown in Figure 2.8. These priorities were the priorities to be presented to proxy patients

during the delivery of the *Expert Choice*-based PDA in clinical consultations. Note from Figure 2.8 that in *Expert Choice* the priorities are normalised to sum to unity.

Figure 2.8. Priorities of the options on the top-level criteria (Expert Choice-based PDA)



3.3.1.3. Summary of STAGE 3 for the *Expert Choice*-based PDA: calculating the priorities of the options on the top-level criteria of the hierarchy.

Having established the three clinical management strategies for the Stage $IIIA_3$ non-small cell lung cancer hypothetical patient in STAGE 1 - neoadjuvant chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2) and best supportive care (A_3)- and the hierarchy of criteria relevant for the decision in STAGE 2 (see Figure 2.4), STAGE 3 of the co-development with clinicians in hospital H1 of the *Expert Choice*-based PDA involved calculating the priorities of the three options on the top-level criteria of the hierarchy. The steps for this calculation involved:

1. Deciding how to calculate the priorities of the options on the bottom-level criteria. It was agreed with the three clinicians that pairwise comparisons of relative

performance between the options would be used to calculate these priorities for the following criteria: disease-related pain, disease-related dyspnoea, disease-related asthenia and disease-related emotional problems. Direct data would be used to calculate the performance levels of the options on all remaining criteria, for which a number of quantitative variables were defined.

2. Calculating the priorities of the options on the bottom-level criteria. This step involved, first, eliciting judgments from each of the three clinicians to estimate the performance levels of the three options on all the bottom-level criteria of the hierarchy. To calculate the performance levels of the options on criterion life expectancy, a Markov model was used, the inputs of which were based on clinicians' judgments. These performance levels were transformed into AHP priorities using the procedures described in Figure 2.5.

3. Finally, the priorities of the options on the bottom-level criteria were propagated up the hierarchy using equal weights for all the children criteria of each top-level criterion to obtain the priorities in Figure 2.8.

The next section describes STAGE 3 of the PDA development process for the *ALEL*-based PDA.

3.3.2. STAGE 3: Calculating the scores of the options on the top-level criteria for the *ALEL*-based PDA.

As with the *Expert Choice*-based PDA, two steps were required using *ALEL* to calculate the scores (from now on termed ratings for consistency with the Annalisa nomenclature) of the three clinical management options on the top-level criteria of the hierarchy. The first step was calculating the ratings of the options on the bottom-level criteria of the hierarchy. The second step was propagating the ratings of the options on the bottom-level criteria up the hierarchy for the *ALEL*-based PDA. These steps are explained below.

3.3.2.1. Calculating the ratings of the options on the bottom-level criteria for the *ALEL*-based PDA.

The calculation of the ratings of the options on the bottom-level criteria required two steps: 1) deciding with the three clinicians in hospital H2 how to measure the ratings, and 2) actually calculating these ratings. The first step is presented in section 2.3.2.1.1. The second step is presented in section 2.3.2.1.2.

3.3.2.1.1. Deciding how to calculate the ratings of the options on the bottom-level criteria for the *ALEL*-based PDA.

To calculate the ratings of the options on the bottom-level criteria with *ALEL*, which implements the SAW MCDA approach, two steps were required: 1) measuring the performance levels of the options on the criteria, 2) transforming these levels into 0-1 ratings on ratio scales. Each option rating on each criterion was then idealised - that is, normalised by the rating of the highest performing option on that criterion.

In a meeting with the three clinicians from hospital H2, the present author proposed and the clinicians agreed to use the same variables measuring the performance levels of the options on the bottom-level criteria that were used for the *Expert Choice*-based PDA in hospital H1 for criteria $C_1(b)$, $C_2(b)$, and $C_7(b)$-$C_{24}(b)$. The descriptions of these variables have already been provided in Table 2.1. For criteria $C_3(b)$-$C_6(b)$, new variables measuring option performance were defined for the *ALEL*-based PDA, as pairwise comparisons between options could not be used. Table 2.3, which is self-explanatory, describes these variables.

Table 2.3. Description of the variables measuring the levels of the options on the criteria $C_3(b)$-$C_6(b)$ (ALEL-based PDA)

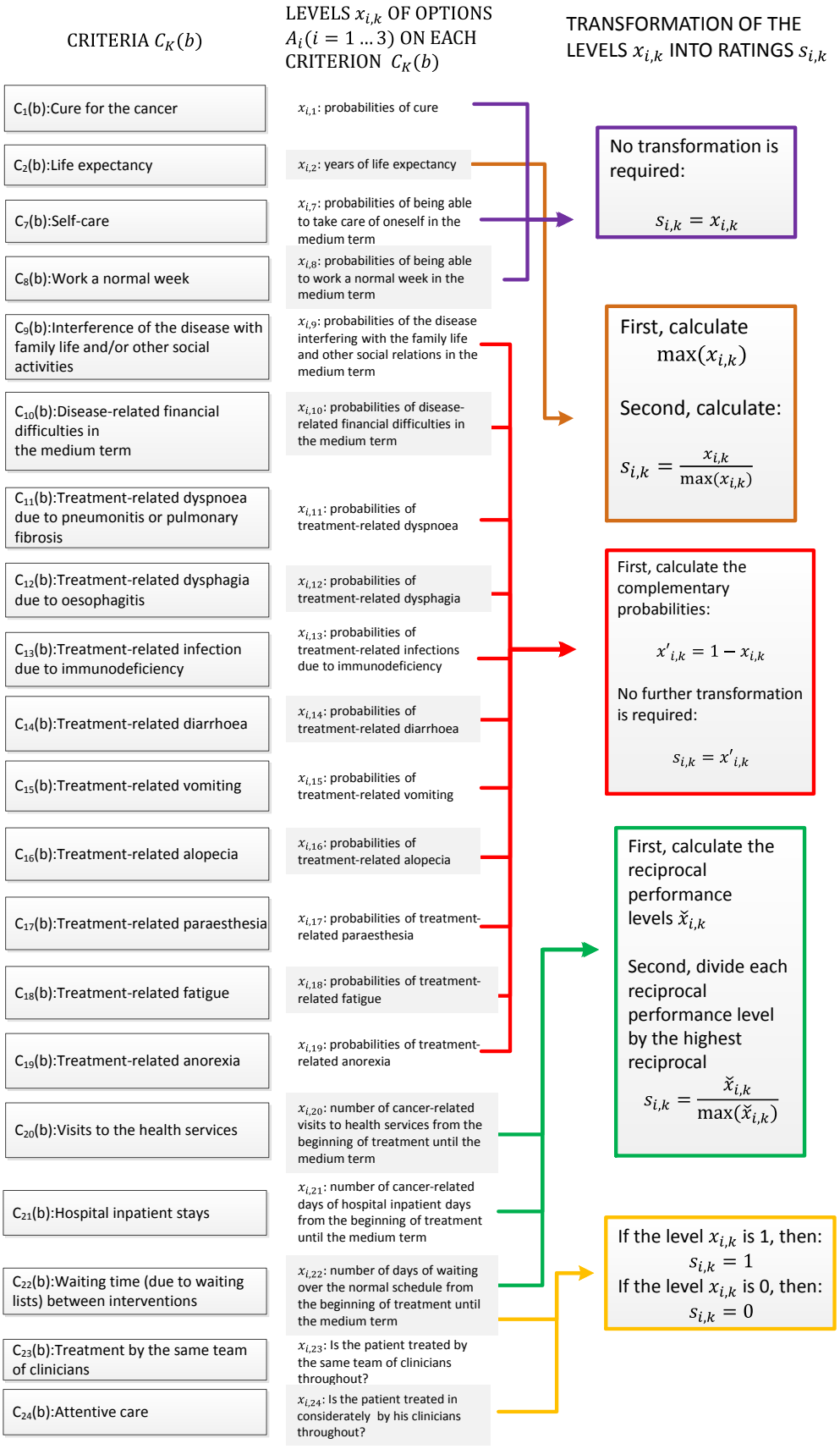
Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \ldots 3)$ on the criterion.**
Disease-related pain $C_3(b)$	Intensity of disease-related pain in the medium term* on a scale between 1 and 9, where: No pain=1 Extreme pain=9	Min: No pain Max: Extreme pain	$x_{i,3} \in \{1,2,3,4,5,6,7,8,9\} \subset \mathbb{N}^+$
Disease-related dyspnoea $C_4(b)$	Intensity of disease-related dyspnoea in the medium term* on a scale between 1 and 9, where: No dyspnoea=1 Extreme dyspnoea=9	Min: No dyspnoea Max: Extreme dyspnoea	$x_{i,4} \in \{1,2,3,4,5,6,7,8,9\} \subset \mathbb{N}^+$
Disease-related asthenia $C_5(b)$	Intensity of disease-related asthenia in the medium term* on a scale between 1 and 9, where: No asthenia=1 Extreme asthenia=9	Min: No asthenia Max: Extreme asthenia	$x_{i,5} \in \{1,2,3,4,5,6,7,8,9\} \subset \mathbb{N}^+$

Table 2.3 (cont.). Description of the variables measuring the levels of the options on the criteria $C_3(b)$-$C_6(b)$ (ALEL-based PDA)

Criterion $C_k(b)$	**Measure of the levels of each option on each criterion**	**Minimum (Min) and maximum (Max) levels of each consequence**	**Variable measuring the levels of the options A_i $(i = 1 \dots 3)$ on the criterion.**
Disease-related emotional problems $C_6(b)$	Intensity of disease-related emotional problems (depression and/or irritability and/or worry) in the medium term* on a scale between 1 and 9, where: No emotional problems=1 Extreme emotional problems=9	Min: No emotional problems Max: Extreme emotional problems	$x_{i,6} \in \{1,2,3,4,5,6,7,8,9\} \subset \mathbb{N}^+$

Once the variables used to measure the performance levels of the options on the bottom-level criteria of the hierarchy were defined, the next step was rating these performance levels. Figure 2.9 describes the procedures used for transforming option performance levels into ratings for criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$. Figure 2.10 describes the procedures used for transforming option performance levels into ratings for criteria $C_k(b)$ $(k = 3,4,5,6)$.

Figure 2.9. Transformation of the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ on criteria $C_k(b)$ $(k = 1,2,7,8 \dots 24)$ into ratings $s_{i,k}$ (ALEL-based PDA)



From Figure 2.9, note five types of transformations, each highlighted with a colour code. Each of these transformations is explained below:

1. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 1,7,8)$, respectively cure for cancer, being able to take care of oneself in the medium term (i.e. two years after the start of treatment), and being able to work a normal week in the medium term (i.e. two years after the start of treatment), colour coded in purple in Figure 2.9. These performance levels were, respectively, the probability of cure, the probability of being able to take care of oneself in the medium term, and the probability of being able to work a normal week in the medium term with each option. Higher probabilities indicated better outcomes. In SAW, these probabilities did not need transformation to be used as ratings, as is illustrated in the purple box in Figure 2.9;

2. Transformation of performance levels $x_{i,k}$ on criterion $C_2(b)$, life expectancy, colour-coded in brown in Figure 2.9. Life expectancy was measured in years. In order to transform years of life expectancy into ratings such that a higher rating reflected higher life expectancy and a lower rating reflected lower life expectancy, this was achieved by dividing the years of life expectancy for each option by the life expectancy of the highest performing option, as shown in the brown box in Figure 2.9;

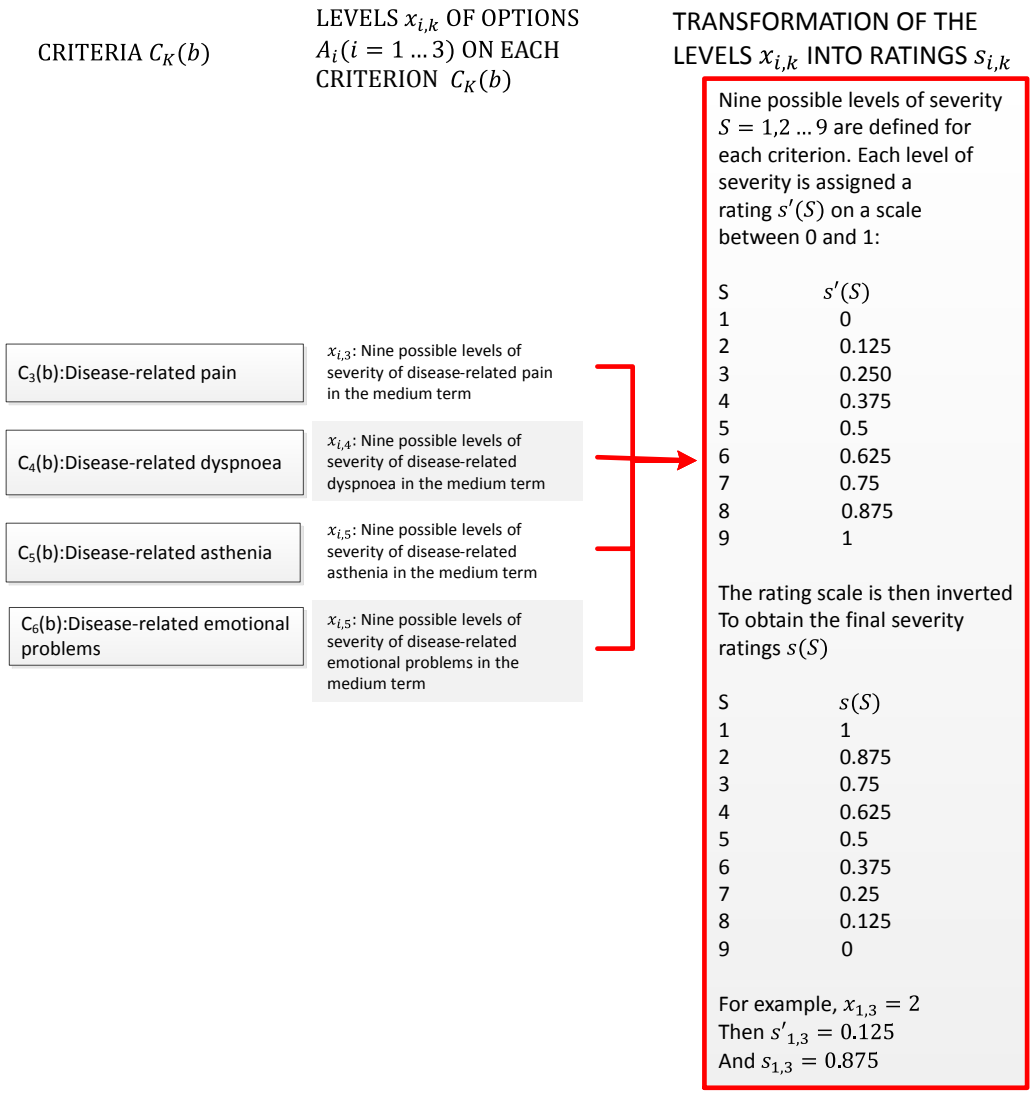
3. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 9 \ldots 19)$, adverse effects of the clinical management options, colour-coded in red in Figure 2.9. These performance levels were all probabilities of adverse effects due to the clinical management options. Their transformation into ratings so that a higher rating indicated a better outcome (i.e. less adverse effects) and a lower rating indicated a worse outcome (i.e. more adverse effects) was done by transforming the probabilities of adverse effects into the probabilities of avoiding adverse effects. This is achieved by calculating the complementary probabilities, as indicated in the red box in Figure 2.9.

4. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 20,21,22)$, visits to the health services, hospital inpatient stays, and waiting time (due to waiting lists) between interventions, colour-coded in green in Figure 2.9. These performance levels were measured respectively as the number of visits to the health services, the number

of days in hospital, and the number of days of waiting time (due to waiting lists) between interventions. The transformation of these performance levels into ratings such that a higher rating reflected a better outcome and a lower rating reflected a worse outcome required two steps. First, the reciprocals of these performance levels were calculated. Second, they were transformed into ratings by dividing the performance level for each option by the performance of the highest performing option on each criterion, as shown in the green box in Figure 2.9;

5. Transformation of performance levels $x_{i,k}$ on criteria $C_k(b)$ $(k = 23{,}24)$, treatment by the same team of clinicians and attentive care (colour-coded in yellow in Figure 2.9). These performance levels were measured by dichotomous 1/0 variables (corresponding to statements “yes”/”no”). The transformation of these performance levels into ratings was done simply by assigning a performance level of 1 a rating of 1 and assigning a performance level of 0 a rating of 0.

Figure 2.10. Transformation of the levels of performance $x_{i,k}$ of the options $A_i(i = 1,2,3)$ into ratings on $C_3(b)$- $C_6(b)$ (ALEL-based PDA)



From Figure 2.10, the performance levels on criteria $C_k(b)$ $(k = 3,4,5,6)$, disease-related pain, disease-related dyspnoea, disease-related asthenia, and disease-related emotional problems were nine possible levels of severity. To calculate the rating for each of the three clinical management strategies on each of these criteria, each option's level on each criterion was mapped on a scale, as shown at the top of the red box in Figure 2.10. The scale was then inverted.

The above transformations completed the procedure agreed with the clinicians in hospital H2 for rating the performance levels of the three clinical management

options on the bottom-level criteria for the *ALEL*-based PDA. The next step was actually calculating these ratings.

3.3.2.1.2. Calculating the ratings of the options on the bottom-level criteria for the *ALEL*-based PDA.

Once the procedure to rate the performance levels of the options – recall, neoadjuvant chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2) and best supportive care (A_3)- on the bottom-level criteria of the hierarchy was agreed with the clinicians in hospital H2, the next step in quantifying the evidence was calculating these ratings. This required two additional steps. First, obtaining estimates for the performance levels $x_{i,k}$ of the options $A_i(1,2,3)$ on each of the bottom-level criteria $C_k(b)$ $(k = 1,2 \dots 24)$. Second, transforming these performance levels into ratings using the procedures outlined in the previous section. The first step is explained below, the second is immediate applying the relevant formulas in Figure 2.9, and will not be further described in the main body of the text.

Considering the time constraints that the three clinicians from hospital H2 confronted, the chosen source of the evidence information for the *ALEL*-based PDA was, as with the *Expert Choice*-based PDA in hospital H1, the clinicians' judgments. These judgments were elicited for each of the three clinicians (the pulmonologist, the oncologist, the thoracic surgeon).

To elicit these judgments, a questionnaire was developed by the present author using the *ELICIA* online survey functionality. The questionnaire, termed the *ALEL-based PDA evidence generation questionnaire* (*ALEL*-EGQ) was composed of 88 questions. It was developed to elicit, from each of the three clinicians in hospital H2, two types of judgments:

- First, judgments about the levels $x_{i,k}$ $(i = 1 \dots 3; k = 1,3,4 \dots 24)$ of the options A_i $(i = 1 \dots 3)$ on the criteria $C_1(b)$, and $C_3(b)$-$C_{24}(b)$;

- Second, matrices of transition probabilities for input into a Markov model to calculate the levels $x_{i,2}$ $(i = 1 \dots 3; k = 2)$ on criterion $C_2(b)$, i.e. the life expectancy of the hypothetical patient with each of the three clinical management strategies;

Each of the three clinicians from hospital H2 was asked to fill in the *ALEL*-EGQ questionnaire online. Three sets of responses were obtained, one set for each clinician.

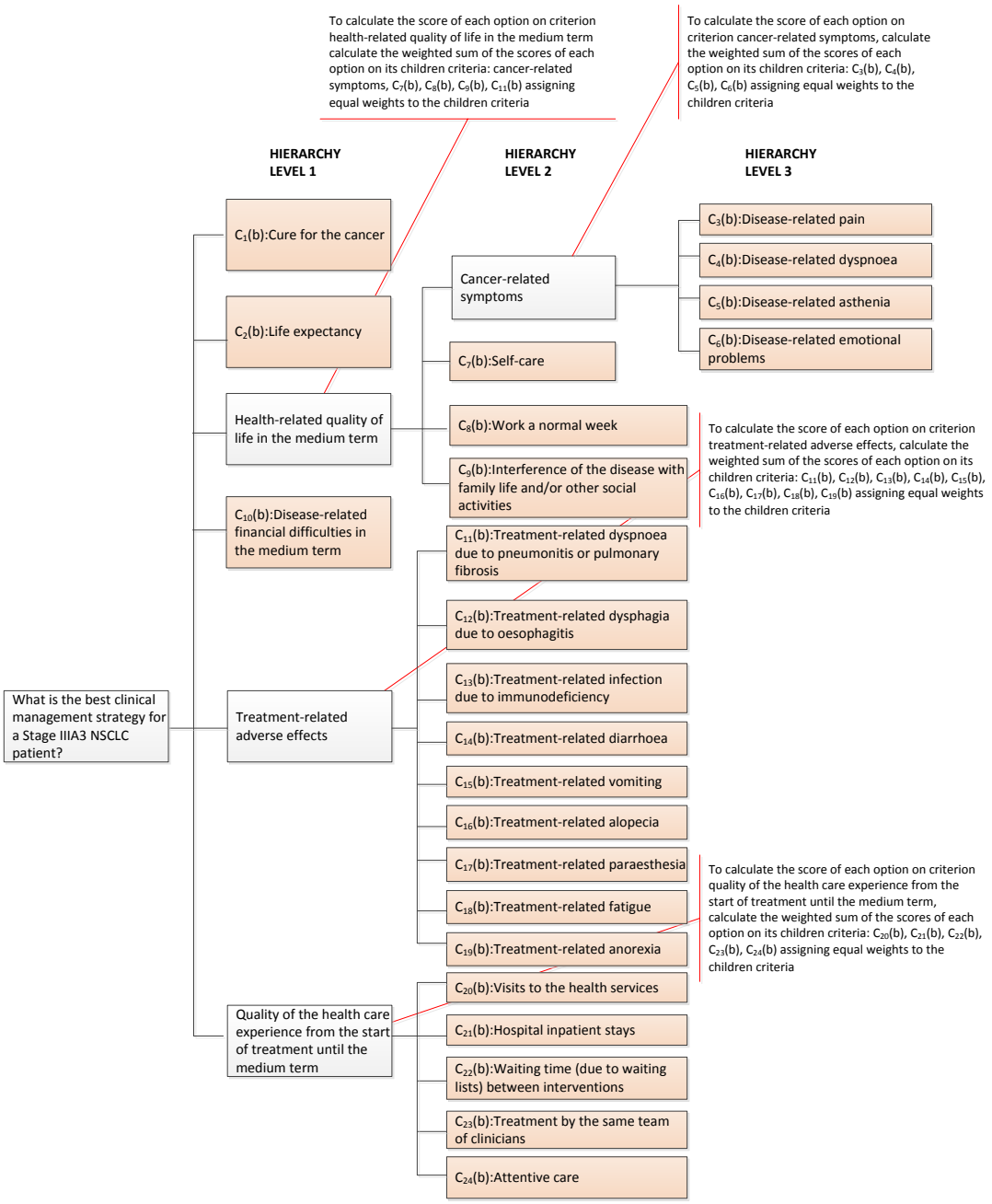
Appendix 5 presents the *ALEL*-EGQ, the judgments elicited from each of the three clinicians, and the corresponding ratings of the options on the bottom-level criteria, calculated as described in Figures 2.9 and 2.10.

The next step was calculating the ratings of the three options on the top-level criteria of the hierarchy using *ALEL*. This procedure is explained in the next section

3.3.2.2. Propagating the ratings of the options on the bottom-level criteria up the hierarchy for the *ALEL*-based PDA.

Figure 2.11 below shows that three of the criteria in the hierarchy were parent criteria with sub-criteria. They are: cancer-related symptoms (on Level 2 of the hierarchy), quality of life in the medium term (on Level 1 of the hierarchy), treatment-related adverse effects, and quality of the health care experience from the start of treatment until the medium term (Level 1 of the hierarchy). *ALEL*, unlike *Expert Choice*, did not automatically calculate the ratings of the options on each parent criterion once weights were assigned to all its children criteria. Instead, a weighted sum had to be calculated for each group of children criteria. This is illustrated in Figure 2.11.

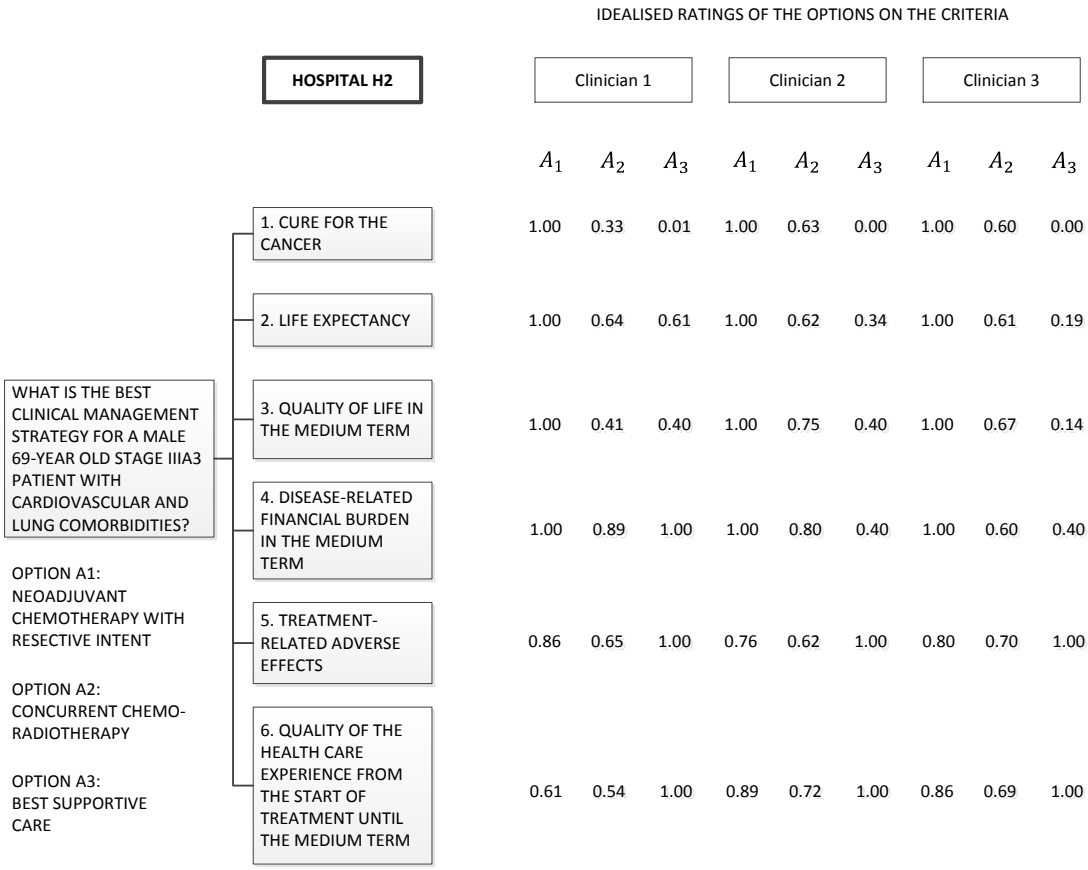
Figure 2.11: Weighted sums required in order to calculate the ratinss of the options on the parent top-level criteria (ALEL-based PDA)



From Figure 2.11, to calculate the rating of each option on a parent criterion, the weighted-sum of the ratings of that option on its children criteria needed to be calculated. Once the weighted-sums were calculated for each group of children criteria, the ratings of the options on the top-level criteria were calculated. These ratings, presented in Figure 2.12, were those to be presented to patients during the

delivery of the *ALEL*-based PDA in clinical consultations. Note from Figure 2.12 that the option ratings were idealised.

Figure 2.12. Calculating the ratings of the options on the top-level criteria (ALEL-based PDA)



3.3.2.3. Summary of STAGE 3 for the *ALEL*-based PDA: calculating the ratings of the options on the top-level criteria of the hierarchy.

Having established the three clinical management strategies for the Stage $IIIA_3$ non-small cell lung cancer hypothetical patient in STAGE 1 - neoadjuvant chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2) and best supportive care (A_3)- and the hierarchy of criteria relevant for the decision in STAGE 2 (see Figure 2.4), STAGE 3 of the co-development with clinicians in hospital H2 of the *ALEL*-based PDA involved calculating the ratings of the three options on the top-level criteria of the hierarchy. The steps for this calculation involved:

1. Deciding how to calculate the ratings of the options on the bottom-level criteria. It was agreed with the three clinicians to define a number of quantitative variables for each bottom-level criterion to measure the performance levels of the options on these variables and then transform these into ratings;

2. Calculating the ratings of the options on the bottom-level criteria. This step involved, first, eliciting judgments from each of the three clinicians to estimate the performance levels of the three options on all the bottom-level criteria of the hierarchy. To calculate the performance levels of the options on criterion life expectancy, a Markov model was used, the inputs of which were based on clinicians' judgments. These performance levels were transformed into ratings using the procedures described in Figure 2.9.

3. Finally, the ratings of the options on the bottom-level criteria were propagated up the hierarchy calculating weighted sums of all the children criteria of each parent criterion as described in Figure 2.12.

The next section provides an analysis of the results of using *Expert Choice* to develop the Stage $IIIA_3$ non-small cell lung cancer PDA in hospital H1 and of using *ALEL* to develop the same PDA in hospital H2.

3.4. Analysis.

3.4.1. Comparing the process of developing the Stage $IIIA_3$ non-small cell lung cancer PDA with clinicians in hospital H1 and in hospital H2.

The process of development of the PDA with clinicians using *Expert Choice* in hospital H1 and using *ALEL* in hospital H2 had both similarities and differences, described for each stage of PDA development below.

3.4.1.1. STAGE 1: determining the relevant options for the clinical management of the hypothetical lung cancer patient.

The process of determining the relevant options for the clinical management of a hypothetical 69 year-old State $IIIA_3$ (TNM stage T2N2M0) non-small cell lung cancer patient was common to both hospital H1 and hospital H2. It was based on this author's review of current clinical guidelines for the treatment of non-small cell lung cancer and adapted by individual contributions of the three clinicians in each

hospital. This led to the same three options in each of the two hospitals: neoadjuvant chemotherapy (option A_1) with resective intent, concurrent chemo-radiotherapy (option A_2) and best supportive care (option A_3). The first two options were composed of active treatments (e.g. chemotherapy, surgery and/or radiotherapy), the third option is palliative. The pathway of interventions (each dependent on an uncertain event, see Figures 2.1 and 2.2) within each of the two active clinical management strategies was also common to both hospitals. This highlights the similarities in the treatment protocols across hospitals.

With respect to the third option, best supportive care, it was defined in both hospitals in broad terms. This is because it is composed of a sequence of on-demand, piecemeal interventions. Each team of three clinicians confirmed that the palliative care protocols applied in their respective hospitals were based on current guidelines. In addition, the author of this study verified that the mode of provision of best supportive care did not differ much across the two hospitals. Hospital Germans Trias I Pujol (hospital H1) in Badalona has a unit of integrated palliative care. The regional government of Catalonia has a programme of home care (Programa de atencio domiciliaria i equips de support, PADES [243]) which is active in the catchment area of hospital H1. Doctors, nurses and social workers from PADES coordinate their activities with the palliative care unit of hospital H1. Hospital Reina Sofia (hospital H2) in Cordoba also has unit of palliative care, which provides home care for patients in need. From the above it is assumed that the definition of best supportive care (option A_3) did not differ substantially across the two hospitals.

3.4.1.2. STAGE 2: determining the criteria relevant for the hypothetical patient for choosing between the three options.

In each of the two hospitals, the three clinicians jointly determined, with guidance from this researcher, the hierarchy of criteria considered relevant for a hypothetical Stage $IIIA_3$ lung cancer patient in making the choice between the three clinical management options. In both cases, the chosen criteria were remarkably similar. In both cases the team of clinicians initially proposed to use the EORTC QDQ-C30 items as the basis for the quality of life sub-criteria. The resulting hierarchy (see Figure 2.3) was modified by this author to 1) avoid using the EORTC QDQ-C30, 2) to minimise its size and 3) in an attempt to minimise the number of criteria which

were not preference independent while respecting the choice of criteria made by the clinicians.

The final hierarchy of criteria had twenty-four bottom-level criteria. In hospital H1, using the *Expert Choice* template as the basis for the lung cancer PDA, the three clinicians considered that although the template allowed for a hierarchical representation of the decision problem to the hypothetical patient the hierarchy was too large to be communicated to the patient during a standard clinical consultation. For this reason, the clinicians decided that only the six top-level criteria would be presented in the PDA. In hospital H2, the hierarchical representation of the decision problem was not a choice, as *ALEL* does not allow for such a hierarchical representation. The implications are clear: lack of time did not allow to take full advantage of the capacity of *Expert Choice* to represent the decision problem in all its richness via the presentation of the full hierarchy of decision criteria.

3.4.1.3. STAGE 3: Calculating the scores of the options on the top-level criteria for the *Expert Choice*-based PDA and the *ALEL*-based PDA.

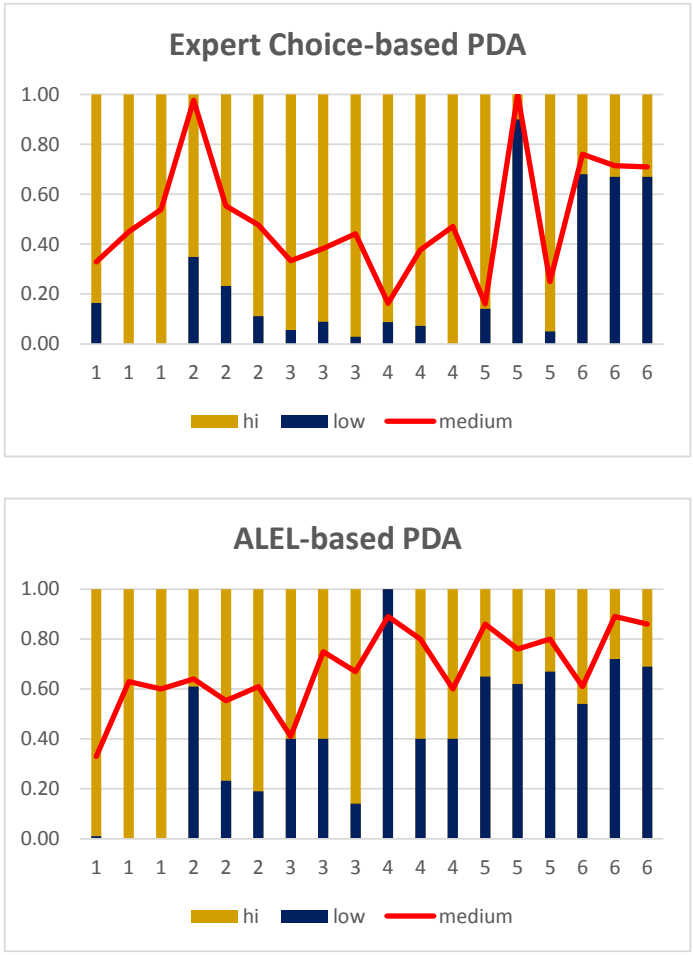
The calculation of the single-criterion scores of the three clinical management options on the six top-level criteria followed in general the same approaches in both hospital H1 and H2: 1) defining and agreeing with the clinicians variables for measuring the performance levels of the options on the bottom-level criteria, 2) measuring the levels of the options on these criteria based on individual clinician judgments, 3) transforming these levels into scores (priotities for the *Expert Choice*-based PDA, ratings for the *ALEL*-based PDA), and 4) propagating these scores up the hierarchy without injecting preferences.

For the *Expert Choice*-based PDA, the calculation of the priorities of the options on twenty of the twenty-four bottom-level criteria used direct data based on clinician judgments instead of pairwise comparisons of relative performance of the options. This was done in order to avoid asking the clinicians in hospital H1 to make an unreasonable (about four hundred) pairwise comparisons. Instead, they were asked to make one hundred and eighteen judgments. For the *ALEL*-based PDA, eighty-eight judgments were required from each of the clinicians in hospital H2 to obtain the ratings of the options on the twenty-four bottom-level criteria. This exercise of expert judgment elicitation was a feasible way to obtain the evidence for input in the model

without having to review the literature, a task for which neither team of clinicians had time.

Figure 2.13 displays in blue/orange stacked bars the lowest/highest six scores corresponding to the lowest/highest performing options on the six top-level criteria according to the judgments of 1) the three clinicians in hospital H1 (for the *Expert Choice* PDA) and of 2) the three clinicians in hospital H2 (for the *ALEL*-based PDA). The red line shows the six scores on the six top-level criteria corresponding to the mid-performing option according to the judgments of 1) the three clinicians in hospital H1 (for the *Expert Choice* PDA) and of 2) the three clinicians in hospital H2 (for the *ALEL*-based PDA).

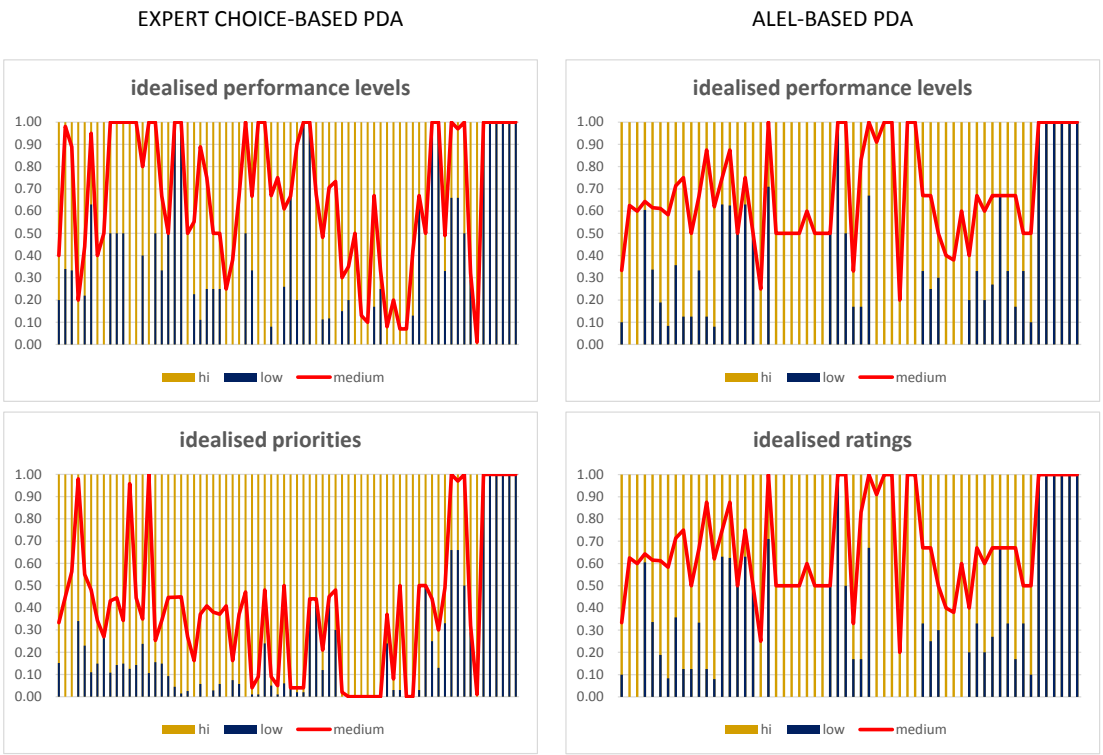
Figure 2.13. Visual representation of the scores (priorities/ratings) of the options on the six top-level criteria



From Figure 2.13, focusing on the red lines in both graphs, there is more variation in the single-criterion scores of the mid-performing options for the *Expert Choice*-based PDA than for the *ALEL*-based PDA. Focusing on the stacked bars, note that the closer the highest and lowest scores of the lowest/highest performing options are for each criterion the more similar the blue and orange bars will be in size. The highest and lowest scores of the lowest/highest performing options are more dissimilar in size for the *Expert Choice*-based PDA than for the *ALEL*-based PDA.

In short, from Figure 2.13, the priorities of the options on the six top-level criteria for the *Expert Choice*-based PDA are more scattered than the ratings of the options on the six top-level criteria for the *ALEL*-based PDA. To get an indication of to what extent this is due to differences in clinicians' judgments and to what extent it is due to the procedures used to transform the performance levels into priorities/ratings, it is useful to compare the performance levels of the options and their scores (i.e. priorities/ratings) for the bottom-level criteria. Figure 2.14 shows this comparison.

Figure 2.14. Comparing the performance levels and the scores (priorities/ratings) of the options on the bottom-level criteria



From Figure 2.14, notice that for the *Expert Choice*-based PDA the transformation of the performance levels into priorities creates a distortion in a number of outcomes, while this is not the case with the *ALEL*-based PDA. This distortion in the *Expert Choice*-based PDA occurs for the criteria where the performance levels of the options are measured as probabilities and which are then transformed into priorities using the odds transformation.

3.4.2. Estimating the reliability of clinicians' judgments for the *Expert Choice*-based PDA (hospital H1) and for the *ALEL*-based PDA (hospital H2).

The inter-rater reliability in the judgments about the performance levels of the clinical management strategies on the bottom-level criteria of the hierarchy was estimated by calculating for the *Expert Choice*-based PDA (hospital H1) and for the *ALEL*-based PDA (hospital H2) the Intraclass Correlation Coefficient (ICC) for each option via a random effects two-way analysis of variance (ANOVA) for a random sample of three clinicians scoring twenty items, as indicated by Schrout and Fleiss [244]. The unit of analysis is a single rater. Table 2.4 shows the results of the analysis. Appendix 10 (Panel 1) shows the results of the analysis, which was undertaken using STATA 14 [245].

Table 2.4. Inter-rater reliability

	Expert Choice-based PDA (H1)			*ALEL*-based PDA (H2)		
Option	ICC	CI*(-)	CI*(+)	ICC	CI*(-)	CI*(+)
A_1	0.67	0.44	0.84	0.68	0.45	0.84
A_2	0.62	0.38	0.81	0.88	0.78	0.95
A_3	0.63	0.38	0.81	0.34	0.08	0.62

*95% Confidence interval

From Table 2.4, the ICC was above 0.6 for the three clinical management strategies in hospital H1 and for two (option A_1 and option A_2) clinical management strategies in hospital H2. The ICC was 0.34 for option A_3 in hospital H2. Following Landis [246] and Fleiss and Cohen [247], in the first case (ICC>0.6) the agreement between clinicians can be considered substantial and in the second case (ICC=0.34) low.

3.4.3. Comparing the scores on the top-level criteria for the *Expert Choice*-based PDA (hospital H1) and for the *ALEL*-based PDA (hospital H2).

In order to identify if, jointly evaluated, there were differences in the average scores (priorities in the *Expert Choice*-based PDA/ratings in the *ALEL*-based PDA) on the top-level criteria by clinician, option, and criterion, multiple linear regression analysis was used. Two regression models were fit, one for each PDA/hospital. The dependent variable was the score (priority/rating). The independent variables were 1) the clinician, 2) the option, and 3) each of the six top-level criteria. Appendix 10 (Panel 2) shows the outputs of the analyses, undertaken using STATA 14. The output shows the coefficients, comparing each level of the independent variables with the reference level. The Wald test was used to verify if there were differences in the scores (priorities/ratings) between levels of the independent variables.

In hospital H1, there was a problem of colinearity. Hence, one option was eliminated from the analysis (option A_3, best supportive care). Including only options A_1 and A_2 in the regression, the above regression model was appropriate to explain the variations in priorities ($F=11.64$, $p=0.0000$). The average priorities differed across options and across criteria, but not across clinicians.

In hospital H2, there was no colinearity. The appropriateness of the above regression model was confirmed ($F=4.13$, $p=0.0007$). The average ratings differed across options, possibly across criteria ($p=0.055$), but not across clinicians).

In other words, the clinicians were consistent in scoring across options and criteria. These results are consistent with the inter-rater agreement observed across clinicians.

3.4.4. Implications for clinical practice.

The case for this research study, a hypothetical 69 year old patient with Stage $IIIA_3$ (TNM stage T2N2M0) non-small cell lung cancer with lung and cardiovascular comorbidities is a good example of a difficult, preference-sensitive decision. For such a patient, the extension of the disease is in the frontier between localised and systemic disease. The frequency of presentation of this type of case in day-to-day clinical practice is low: lung cancer nowadays is usually diagnosed at earlier stages. This circumstance is likely to be reflected in the limited scientific evidence that this researcher found regarding the performance of the three clinical management options

on the large number of specific criteria considered by the clinicians relevant for the decision. The lack of evidence highlights the role of clinical expertise (i.e. expert opinion) in the provision of information for this type of decision model. The finding that there was considerable agreement between clinicians in the elicitation of judgments in this sense is encouraging.

The clinicians participating in this study had never been involved in the development of a PDA and they had never used MCDA as the basis of a PDA. They greatly valued how MCDA involved the explicit and systematic expression of all the relevant elements of a complex decision. For example, the definition of the uncertain pathways of interventions of which each active clinical management strategy was composed. This raised the issue of how to best represent these clinical management strategies in the PDAs. The choice of communicating the options in broad terms was pragmatic, motivated overall by the time constraints of clinicians. This aspect of the development of the *Expert Choice*-based and the *ALEL*-based PDAs highlights the importance of taking uncertainty into account in the development of PDAs for the clinical management of Stage $IIIA_3$ lung cancer.

The clinicians also valued greatly the requirement in MCDA of developing a hierarchy of criteria which considered all the aspects of importance to the hypothetical lung cancer patient in making the choice of clinical management strategy. The two PDAs included clinical outcomes of the interventions, but also outcomes related with the quality of care, and outcomes related with the patient's broader quality of life (e.g. being able to work a normal week, or interference of the disease with the patient's family life and other social activities). By explicitly taking into account all of these aspects, the *Expert Choice*-based PDA and the *ALEL*-based PDA provided an accurate and explicit representation of the overall implications of the different clinical management strategies for the patient. This is important for such a condition as lung cancer where the condition and the treatments impact deeply on many aspects of the patient's life.

The strong time constraints confronted by clinicians in the development of the *Expert Choice*-based PDA and the *ALEL*-based PDA meant that the input of this researcher in the development of both PDAs was large. This highlights the role of external support to clinicians in the development of these tools. In this sense, the presence of

a decision analyst to aid clinicians in the development of MCDA-based PDAs for complex decisions such as the clinical management of lung cancer is desirable. Both hospital H1 and hospital H2 are tertiary hospitals from the Spanish NHS. In these hospitals, there exist structures that can provide additional support for the development of such PDAs. For example, hospital tumour committees. These committees are multidisciplinary groups of clinicians which meet regularly with, among other tasks, the task of refining existing cancer clinical care protocols. The outputs of these committees (e.g. clinical practice guidelines) are disseminated rapidly and generally accepted by clinicians from other hospitals.

Among other constraints, the lack of awareness in clinicians regarding PDAs, their lack of time, and the lack of financial compensation are important barriers for the development of these tools by clinicians as part of their day-to-day workload. This researcher considers that the feasibility of this task would be greatly increased with support from the health system. In the Spanish NHS there are research structures with the capacity to provide such support. For example, the Thematic Networks for Cooperative Research (RETICS [248]) or the Centres for Networked Biomedical Research (CIBER [249]), one of which is dedicated to diseases of the lung.

Chapter 3: Using *Expert Choice* and *Annalisa in Elicia* to deliver, with clinicians in the Spanish NHS, a patient decision aid for the clinical management of Stage IIIA3 non-small cell lung cancer in an environment replicating actual clinical consultations.

Chapter 2 described the methods and results related with Objective 1 of this thesis: assessing the use with clinicians in the Spanish NHS of two alternative software applications which support dissimilar MCDA techniques in the development of a PDA in routine clinical practice. Two software applications (i.e. templates) were used to develop a PDA for the choice of best clinical management strategy in Stage $IIIA_3$ non-small cell lung cancer: *Expert Choice* (with three clinicians in hospital H1), and *Annalisa in Elicia* (with three clinicians in hospital H2). *Expert Choice* supports the AHP MCDA approach and *Annalisa in Elicia* (*ALEL*) supports the SAW MCDA approach. The process of development the *Expert Choice*-based PDA and the *ALEL*-based PDA resulted in:

1. Three options for inclusion in each version of the lung cancer PDA: neoadjuvant chemotherapy with resective intent (option A_1), concurrent chemo-radiotherapy (option A_2), and 3) best supportive care (option A_3);

2. A common set of six top-level decision criteria for inclusion in both versions of the lung cancer clinical management PDA: 1) cure from the cancer, 2) life expectancy, 3) quality of life in the medium term, 4) disease-related financial burden in the medium term, 5) treatment-related adverse effects, 6) quality of the health care experience between the start of treatment and the medium term;

3. Three different sets of scores for inclusion in each version of the PDA reflecting the consequences of the options on the six decision criteria. Each set of priorities for the *Expert Choice*-based PDA was based on the judgments of each participating clinician (one pulmonologist, one oncologist, one thoracic surgeon) from hospital H1. Each set of ratings for the *ALEL*-based PDA was based on the judgments of each participating clinician (one pulmonologist, one oncologist, one thoracic surgeon) from hospital H2

This chapter explains the methods and results related to Objective 2 of this thesis, i.e.:

- To assess the use with clinicians in the Spanish NHS of the two alternative MCDA software applications in the implementation of a PDA in an environment replicating actual clinical consultations.

The chapter begins with an overview of the methods used to implement, i.e. deliver, the two PDAs in an environment replicating actual clinical consultations. It then describes the process of delivering the *Expert Choice*-based PDA in hospital H1. This is followed by a description of the process of delivering the *ALEL*-based PDA in hospital H2. The results and the corresponding commentary are presented at the end of the chapter.

1. Overview of methods.

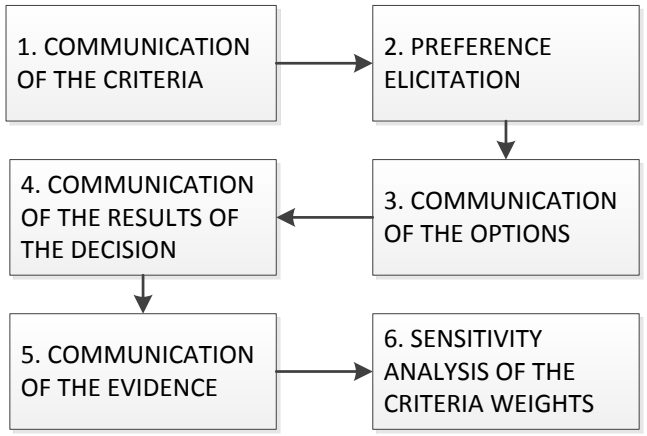
The decision of implementing, i.e. delivering, the *Expert Choice*-based PDA and the *ALEL*-based PDA in an environment replicating actual clinical consultations was motivated by the shared decision-making (SDM) paradigm. SDM can be conceptualised as a continuum with patient-led decision making at one end and physician-led decision making at the other end [250]. In patient-led decision making the clinician provides expert knowledge to the patient but makes no recommendations about the choice [250]. Based on patient-led decision making, the delivery of the *Expert Choice*-based PDA in hospital H1 and the delivery of the *ALEL*-based PDA in hospital H2 were structured by setting up in each of the two hospitals a number of hypothetical consultations. Each hypothetical consultation simulated an actual consultation between 1) a pulmonologist and 2) a patient with Stage $IIIA_3$ (TNM stage T2N2M0) non-small cell lung cancer and lung and cardiovascular comorbidities in which the pulmonologist guides the patient through the PDA but the patient makes the choice of clinical management strategy.

To deliver the *Expert Choice*-based PDA, three hypothetical consultations were set up in hospital H1. In each consultation, one of the three clinicians - the pulmonologist (clinician 1), the oncologist (clinician 2) and the thoracic surgeon (clinician 3) individually guided, with support from the author of this study, a proxy patient (a non-clinical member of hospital staff) through the PDA. Similarly, to

deliver the *ALEL*-based PDA, three hospital consultations were set up in hospital H2 between each individual clinician and a proxy patient. At the end of each hypothetical consultation in both hospital H1 and hospital H2, both the clinician and the proxy patient were asked to assess the quality of the decision using the "My Decision Quality" (MDQ) instrument [21]. The process of delivery of both the *Expert Choice*-based PDA and the *ALEL*-based PDA in the hypothetical consultations was documented.

In the process of delivery of the *Expert Choice*-based PDA (in hospital 1) and the *ALEL*-based PDA (in hospital H2) to the proxy patient, the author of this research identified the following steps (not necessarily in the order presented): 1) communication of the criteria, 2) communication of the options, 3) communication of the evidence (i.e. communication of the scores of the options on each of the criteria), 4) preference elicitation (i.e. criteria weight elicitation), 5) communication of the results of the decision to the patient, and 6) sensitivity analysis of the criteria weights. To deliver the two PDAs, the first decision confronted by the clinicians and the present author was the choice of desired sequence of delivery steps. After group discussions with both teams of clinicians, in both teams it was agreed that the following sequence was appropriate (see Figure 3.1):

Figure 3.1. Sequence of delivery steps of the Expert Choice-based and ALEL-based PDAs



From Figure 3.1., note that the first step in the delivery of either PDA in the hypothetical consultation was the communication of the criteria to the proxy patient, immediately followed by the elicitation of proxy patient preferences (i.e. the elicitation of criteria weights). Next came the communication of the options, followed by the communication of the results of the decision, the communication of the evidence and the sensitivity analysis of criteria weights.

Eliciting proxy patient preferences before communicating information about the options is justified to 1) help the proxy patient focus on the different aspects of importance to him before focusing on any other aspect of the decision, and to 2) reduce the chance of bias in the elicitation of the proxy patient's preferences due to knowledge about the options. Communicating the results of the decision before communicating the evidence was felt by the clinicians as a good way for the proxy patient to understand this evidence. This is because communicating the results before the evidence allows the proxy patient to ask the question of how these results came about, focusing attention on how the options perform on the different criteria. Sensitivity analysis on criteria weights then allowed to focus on the imprecision of the weight estimates during the elicitation of the proxy patient's preferences.

Section 2 explains the steps of delivering the *Expert Choice*-based PDA in hypothetical consultations in hospital H1. Section 3 explains the steps of delivering the *ALEL*-based PDA in hypothetical consultations in hospital H2.

2. Delivering the *Expert Choice*-based PDA in hypothetical clinical consultations in hospital H1.

A group meeting with the clinicians in hospital H1 was set up to decide how to implement the agreed sequence of delivery steps in a hypothetical consultation using the *Expert Choice*-based PDA. In this meeting, it was also agreed that each clinician would guide the proxy patient through the PDA. In the case of *Expert Choice*, this guidance involved helping the proxy patient move through the different *Expert Choice* menus to be able to access each screen relevant for each particular delivery step. It was agreed that the proxy patient could ask questions at any point during the delivery of the PDA.

2.1. Communication of the criteria

Recall that the criteria to be presented in the *Expert Choice*-based PDA were the six top-level criteria of the agreed hierarchy (see Figure 2.3): 1) cure from cancer, 2) life expectancy, 3) quality of life in the medium term, 4) disease-related financial burden in the medium term, 5) treatment-related adverse effects, and 6) quality of the health care experience from the start of treatment until the medium term. The team of clinicians in hospital H1 suggested that the six criteria should be communicated in the PDA using the easiest language possible. Table 3.1 shows the language agreed between the three clinicians and the author of this study to communicate the criteria to the proxy patient.

Table 3.1. Communication of the criteria to the proxy patient (Expert Choice-based PDA)[3]

Criterion	**Language used for communication to the patient**
Cure	To get cured
Life Expectancy	To live longer, independently of my state of health
Quality of life in the medium term*	To have a good quality of life two years from now. I will have a good quality of life two years from now if, after two years: 1) I do not have the following symptoms of cancer: pain, shortness of breath, loss of weight and extreme tiredness, and emotional problems such as depression and irritability 2) I am able to take care of myself without help from others, I can work a normal forty-hour week, and my condition does not interfere with my family life and other social relations
Disease-related financial burden in the medium term*	To avoid having financial difficulties due to my condition two years from now
Treatment-related adverse effects	To avoid the adverse effects due to the treatment. If I take the treatment I can have the following adverse effects: breathlessness, problems swallowing, infections due to a drop in my defences, diarrhoea and vomiting, loss of hair, a prickling or burning sensation (especially in the hands and feet), fatigue, and loss of weight. All of these adverse effects with the exception of the loss of hair can be serious enough to require that I am admitted to hospital
Quality of the health care experience from the start of treatment until the medium term*	To have a good experience of the health care received in the next two years. I will have a good experience of the health care received if 1) I have to go to the outpatient clinics as little as possible, if 2) I have to be admitted to hospital as few times as possible, 3) I do not have to wait unnecessarily to receive treatment, 4) I am always treated by the same clinician or clinicians, 5) I am treated in an attentive and considerate manner by my clinicians

[3] See Appendix 8 for the Spanish criteria names

In each hypothetical consultation, the communication of the criteria to the proxy patient was made by letting the patient read through the criteria in the *Tree View* screen of *Expert Choice* (see Figure A6.1 in Appendix 6 for an illustration). Once the proxy patient considered that he had understood the criteria, the clinician helped the patient to access the screen for preference elicitation.

2.2. Preference elicitation.

Proxy patient preferences for the different criteria were elicited using pairwise comparisons of relative importance between the criteria. For these pairwise comparisons, the standard 1-9 AHP scale was used (see Table 3.2). The clinicians were given the option of using either numbers, verbal expressions or sliding bars to elicit proxy patient preferences in *Expert Choice*. They all agreed to use sliding bars because they thought the sliding bars were visually intuitive to express the pairwise comparisons (see Figure A6.2 in Appendix 6 for an illustration).

Table 3.2. 1-9 AHP scale of intensity of dominance.

Intensity of dominance	Definition
1	Equal intensity of dominance
2	Weak intensity of dominance
3	Moderate intensity of dominance
4	Moderate plus intensity of dominance
5	Strong intensity of dominance
6	Strong-plus intensity of dominance
7	Very strong intensity of dominance
8	Very, very strong intensity of dominance
9	Extreme intensity of dominance

The preference elicitation was conducted in each hypothetical consultation as follows. First, the clinician explained to the proxy patient how to perform a pairwise comparison of relative importance between criteria. Each proxy patient had to make fifteen of these pairwise comparisons. Any inconsistencies in the proxy patient's judgements were highlighted in real time in *Expert Choice* by a visual display of the

inconsistency index which was monitored by the clinician. A consistency index higher than the (more or less arbitrary) value of 0.1 required a reassessment of the pair-wise comparisons. Interestingly, in hospital H1 the pulmonologist raised the issue that perhaps the threshold of the consistency index should be increased from 0.1 to 0.2. The reason for this was, according to the pulmonologist, that expecting high consistency in the pair-wise comparison judgments of a real lung cancer patient was perhaps expecting too much. For this clinician's hypothetical consultation, the consistency index threshold was raised to 0.15 to accommodate his views.

Once the criteria weights were obtained, the clinician helped the patient to access the screen for the communication of the options.

2.3. Communication of the options.

This step of PDA delivery involved describing to the proxy patient the three clinical management options, i.e. neoadjuvant chemotherapy with resective intent (option A_1), concurrent chemo-radiotherapy (option A_2), and best supportive care (option A_3). This was a challenging task, as the actual sequence of interventions within each strategy is uncertain (see Figures 2.1 and 2.2 in Chapter 2). The clinicians and this researcher agreed to describe the options in terms of a broad outline of the interventions that could occur within each option.

Table 3.3 shows the textual content of the three descriptions. Note that each description is in the first person, for consistency with the textual description of the criteria shown in Table 3.1. Each option was communicated to the proxy patient by letting the patient read the textual content from Table 3.3 for each option using *Expert Choice*'s information document screen (for an illustration, see Figure A6.3 in Appendix 6).

Table 3.3. Textual content of the three option descriptions (Expert Choice-based PDA)[4]

Option	**Description**
Neoadjuvant chemotherapy with resective intent (Option A_1)	1. Neoadjuvant chemotherapy with resective intent. This treatment consists in injecting me a medicine (chemotherapy) that kills the cancer cells. The goal is to reduce the size and extent of the tumour enough that it can be taken out using surgery. This treatment has several steps. First, I am given the chemotherapy in several sessions over a month or so. If, after a scan, it looks like the chemotherapy is successful at reducing the size and extent of the tumour so that it can be removed, then I will have a small chest intervention called mediastinoscopy to confirm that this is the case. If the mediastinoscopy confirms that this is indeed the case, then I will have the surgery to remove the tumour. The surgery consists in removing the part of the lung where the tumour is lodged. If the chemotherapy is not successful at reducing the size and extent of the tumour, I will not have surgery. Instead, I will be given chemotherapy along with radiotherapy. Radiotherapy consists in using a machine that sends rays to burn the cells of the tumour. If after either of these treatments the tumour comes back, which is likely, I will be given chemotherapy again;
Concurrent chemotherapy (Option A_2)	2. Concurrent chemo-radiotherapy. This treatment consists in giving me alternatively chemotherapy and radiotherapy for several weeks with the goal of eliminating the tumour. If after the treatment the tumour comes back, I will be given chemotherapy again;
Best supportive care (Option A_3)	3. Best supportive care. This treatment is not directed at curing the tumour, but at eliminating the symptoms of the cancer. I will be given, if and when I require it, the following treatments: physiotherapy, psychotherapy, treatment against pain, treatment against breathlessness, treatment against loss of weight, etc.

[4] See Appendix 8 for the Spanish option names

Once the proxy patient considered that he had understood the options, the clinician helped the patient to access the screen for communicating the results of the decision.

2.4. Communication of the results of the decision

This stage in the delivery of the *Expert Choice*-based PDA involved communicating to the proxy patient the three overall scores of the three options and their ranking from best to worst. The communication of the results of the decision was done using the *Synthesise* screen in *Expert Choice* (see Figure A6.4 in Appendix 6 for an illustration). This screen consists of three horizontal bars reflecting the overall priorities (normalised to sum to one) of the three options. In the Synthesise screen, the options are ranked from best to worst from the top-down. Once the proxy patient considered that he had understood the overall scores of the options and the ranking of the options, the clinician asked the proxy patient if he wanted to see the evidence (i.e. the priorities) of the options on the individual criteria. If the answer was "yes", the clinician helped the patient to access the screen for visualising the evidence. If the answer was "no", the clinician moved to the sensitivity analysis screen. In the three hypothetical consultations in hospital H1 all three proxy patients wanted to see the evidence.

2.5. Communication of the evidence.

This stage in the delivery of the *Expert Choice*-based PDA involved showing the proxy patient the priorities (normalised to sum to one) of the three clinical management strategies on each of the six decision criteria.

The communication of the evidence was done using the *criterion* window in *Expert Choice* (see Figure A6.5 in Appendix 6 for an illustration). The clinician explained to the proxy patient that the priorities of the options on each criterion reflected how much better each option was with respect to each other option in terms of achieving that criterion. Once the proxy patient considered that he had understood the evidence, the clinician helped him to access the sensitivity analysis screen.

2.6. Sensitivity analysis of the criteria weights.

The final stage in the delivery of the *Expert Choice*-based PDAs was the sensitivity analysis on the criteria weights elicited from the proxy patient. The sensitivity analysis assessed the robustness of the overall results (in terms of the ranking of the options) to changes in the criteria weights. One approach to such a sensitivity analysis involves varying the criteria weights over a range. This was the approach used in each hypothetical consultation. Each of the six criteria weights was varied over a range of plus/minus 20%.

To undertake the sensitivity analysis, the three clinicians chose to use the *dynamic sensitivity analysis* screen (see Figure A6.6 in Appendix 6), as they felt it was easy to use for patients. This screen displays on the left side the weights of the decision criteria as sliding horizontal bars, and on the right side the aggregate scores of the three options also as horizontal bars. As each criterion weight is changed, the overall scores of the options change correspondingly. The sensitivity analysis was done for each criterion by the clinician and communicated verbally to the proxy patient. If there was a change in the ranking of the options as a result of any of these weight changes, the patient was asked to state whether or not he felt that the weight of that criterion could vary over the range that induced a change in the ranking of the options.

After the sensitivity analysis, the delivery of the *Expert Choice*-based PDA ended. The clinician and the proxy patient then proceeded to answer to the MDQ decision quality questionnaire.

2.7. Summary of delivery steps for the *Expert Choice* PDA.

To deliver the *Expert Choice*-based Stage $IIIA_3$ non-small cell lung cancer PDA in hospital H1, three hypothetical consultations were set up. In each hypothetical consultation, one of the three clinicians the pulmonologist (clinician 1), the oncologist (clinician 2), and the thoracic surgeon (clinician 2) guided a proxy patient (a non-clinical member of hospital staff), with support from the author of this thesis, through the following PDA delivery steps: 1) communication of the criteria (in simple language) using the template's TreeView screen, 2) criteria weight elicitation (including the verification of the consistency of the proxy patient's judgments) using

the graphical pairwise comparisons screen and a 1-9 intensity of dominance scale, 3) communication of the options (describing each of the options in terms of a broad outline) with the template's information document screen, 4) communication of the results using the template's graphical Synthesise screen, 5) communication (if desired by the proxy patient) of the priorities of the options on each of the criteria on a separate criterion window, and 6) sensitivity analysis of the proxy patient's weights varying each weight over a ±20% using the dynamic sensitivity screen.

3. Delivering the *ALEL*-based PDA in hypothetical clinical consultations

As with the *Expert Choice*-based PDA, a group meeting was organised with the clinicians in hospital H2 to decide how to implement the agreed sequence of delivery steps (see Figure 3.1) using the *ALEL*-based PDA in a hypothetical consultation. In this meeting, it was also agreed that each clinician would guide the proxy patient through the PDA. *ALEL* allows to tailor the sequence of screens to be presented to the proxy patient. There are two types of screens in *ALEL*: 1) the Elicia screens and 2) the Annalisa topic screens. The Elicia screens both present information and elicit information from the user. The Annalisa topic screens constitute the decision making interface of *ALEL*. A typical Annalisa topic screen is structured in three panels: 1) the ratings panel at the bottom of the screen, which includes the names of the options and a visual representation (in horizontal sliding bars) of the rating of each option on each criterion, 2) the weights panel in the middle of the screen including the names of the criteria and a visual representation in horizontal sliding bars of the weights assigned by the patient to each of the criteria, and 3) the scores panel at the top of the screen, including a visual representation of the overall scores of the options. The ratings/weights/scores panels can be hidden from view at any point. For example, when eliciting criteria weights using the horizontal sliding bars, the ratings panel and the scores panel can be hidden. The *ALEL*-based PDA was purposefully built by this researcher using a sequence of Elicia screens and Annalisa topic screens guiding the proxy patient through the PDA delivery steps. It was agreed with the clinicians that the proxy patient could ask questions at any point of the PDA delivery.

3.1. Communication of the criteria

As with the *Expert Choice*-based PDA, the criteria presented in the *ALEL*-based PDA were: 1) cure from cancer, 2) life expectancy, 3) quality of life in the medium term, 4) disease-related financial burden in the medium term, 5) treatment-related adverse effects, and 6) quality of the health care experience from the start of treatment until the medium term. The clinicians in hospital H2 were of the opinion that the criteria should be communicated to the proxy patient in the easiest way possible. The chosen textual content of the criteria communication screen was similar as for the *Expert Choice*-based PDA (see Table 3.1), but expressed in second person singular instead of in first person singular. For example, for criterion quality of life in the medium term (third row in Table 3.1), instead of displaying "I will have a good quality of life in the medium term if…" as in the *Expert Choice*-based PDA, the *ALEL*-based PDA displayed "you will have a good quality of life in the medium term if…".

In each hypothetical consultation, the communication of the criteria was made to the proxy patient using an Elicia screen which was accessible from the welcome screen of the PDA (see Figure A7.1 in Appendix 7). Once the proxy patient considered that he had understood the criteria, he moved to the preference elicitation screen by clicking on a button in the Elicia screen.

3.2. Preference elicitation.

The preference elicitation (i.e. the elicitation of criteria weights) for the *ALEL*-based PDA was undertaken in each hypothetical consultation using a two-step process. First, the proxy patient was prompted by an Elicia screen to assess the relative importance of the six criteria on five verbal levels of importance (see Table 3.4, first column). As shown in Table 3.4, each level of importance was assigned a weight between 0 and 1. For an illustration of the screen used to elicit these weights see Figure A7.2 in Appendix 7.

Table 3.4. Levels of criterion importance and associated weights (ALEL-based PDA)

Levels of importance	**Weights**
Not important at all	$w(\text{not important at all}) = 0$
A little bit important	$w(\text{a little bit important}) = 0.25$
Moderately important	$w(\text{moderately important}) = 0.5$
Very important	$w(\text{very important}) = 0.75$
Extremely important	$w(\text{extremely important}) = 1$

Second, upon clicking on a button in the Elicia screen the proxy patient was informed that in the next screen he would visualise in sliding bars the weights corresponding to the verbal levels of importance assigned to each of the criteria. He was advised to modify these weights using the sliding bars until they represented adequately his relative preferences for the different criteria (see Figure A7.3 in Appendix 7). Upon clicking on a button in the Elicia screen the proxy patient was directed to an Annalisa topic screen where he could change the weights that he had previously assigned verbally if he so wished. The weighting panel in the Annalisa topic screen is unique in that when the cursor is dragged across each criterion name a pop-up appears with information about the criterion. Advantage was taken of this feature to include information about each criterion in the weightings panel. Table 3.5 provides a description of the information included in the criteria pop-ups.

Table 3.5. Content of the criterion information pop-ups (ALEL-based PDA)

Criterion (SHORT NAME)	**Content of the pop-up**
Cure (CURE)	Shows how important it is for you to get cured
Life expectancy (TIME ALIVE)	Shows how important it is for you to live longer, independently of the state of health you are in
Quality of life in the medium term (QUAL.LIFE)	Shows how important it is for you to have a good quality of life two years from now
Financial burden due to the disease in the medium term (FIN.DIFF)	Shows how important it is for you to avoid the financial difficulties due to the disease two years from now
Treatment-related adverse effects (ADV.EFF)	Shows how important it is for you to avoid the treatment-related adverse effects
Quality of the health care experience (QUAL.CARE)	Shows how important it is for you to have a good experience of the health care received during the next two years

Once the proxy patient assigned weights to the criteria, the PDA prompted him to move to the options communication screen by pressing a button.

3.3. Communication of the options.

The communication of the options for the *ALEL*-based PDA was done in a similar fashion as for the *Expert Choice*-based PDA. The textual content of the option descriptions was almost identical to that of Table 3.3, except that it was expressed in the second person singular. The information was presented to the proxy patient using an Elicia screen (see A7.4 in Appendix 7 for illustration). Once the proxy patient considered that he understood the options, he moved to the results communication screen by clicking on a button.

3.4. Communication of the results of the decision

This stage in the delivery of the *ALEL*-based PDA involved explaining to the proxy patient the three overall scores of the three options and their ranking from best to worst. The communication of the results of the decision was done using an Annalisa topic screen displaying the overall scores panel and the weightings panel (see Figure A7.5 in Appendix 7 for an illustration). The scores panel consists of three horizontal bars reflecting the overall scores of the three options. Once the proxy patient considered that he understood the overall scores of the options and the ranking of the options, the clinician asked the proxy patient if he wanted to see the evidence of the consequences of the options on the criteria. If the answer was "yes", the patient moved to the evidence communication screen by clicking on a button. If the answer was "no", the clinician stepped in and clicked ahead to the sensitivity analysis screen.

3.5. Communication of the evidence

This stage in the delivery of the *Expert Choice*-based PDA involved showing the proxy patient the ratings of the three clinical management strategies on the six decision criteria. In *ALEL* it was possible to simultaneously visualise all the evidence (i.e. all the ratings of the three options on all the criteria) on the screen. The communication of the evidence was done using the ratings panel of an Annalisa topic screen (see Figure A7.6 in Appendix 7). Once the proxy patient considered that he had understood the evidence, he moved to the sensitivity analysis screen by clicking on a button.

3.6. Sensitivity analysis of the criteria weights

The final stage in the delivery of the *ALEL*-based PDAs was the sensitivity analysis on the criteria weights elicited from the proxy patient. *ALEL* allows for performing sensitivity analyses on the criteria weights within the same Annalisa topic screen as long as the screen is displaying both the weights and the scores panel. In order to perform this sensitivity analysis, the weights assigned by the proxy patient to each of the different criteria were modified by the clinician by an amount of plus/minus 20%. If the ranking of the options changed as a result of a particular criterion weight change, the proxy patient was asked by the clinician to reconsider that criterion weight.

After the sensitivity analysis, the delivery of the *Expert Choice*-based PDA ended. The clinician and the proxy patient then proceeded to answer to the MDQ decision quality questionnaire.

3.7. Summary of delivery steps for the *ALEL* PDA.

To deliver the *ALEL*-based Stage $IIIA_3$ non-small cell lung cancer PDA in hospital H2, three hypothetical consultations were set up. In each hypothetical consultation, one of the three clinicians - the pulmonologist (clinician 1), the oncologist (clinician 2), the thoracic surgeon (clinician 3) guided a proxy patient (a non-clinical member of hospital staff), with support from the author of this thesis, through the following PDA delivery steps: 1) communication of the criteria (in simple language) using an Elicia screen, 2) criteria weight elicitation using first verbal statements and then using an Annalisa topic screen displaying only the weights panel, 3) communication of the options (describing each of the options in terms of a broad outline) with an Elicia screen, 4) communication of the results using an Annalisa topic screen displaying only the overall scores panel and the weights panel, 5) communication (if desired by the proxy patient) of the ratings of the options on all the criteria on the same Annalisa topic screen of the results communication step but now displaying the ratings panel, and 6) sensitivity analysis of the proxy patient's weights varying each weight over a range of ±20%.

The next section describes the use of the "My Decision Quality" (MDQ) after each hypothetical consultation to assess each clinician's and proxy patient's perception of the quality of decision-making using the template-based PDAs.

4. Assessing decision quality using the "My Decision Quality" (MDQ) tool.

The MDQ instrument [21], currently implemented using the *ALEL* template, combines 1) a decision maker's ratings on eight decision-making quality criteria with 2) the weights assigned by the decision maker to these criteria to calculate a score measuring the quality of the subject's decision-making. The eight MDQ criteria are, from the perspective of the decision maker, the following [21]:

1) Being clear about the possible options for him/her and what they involve;

2) Being clear about the possible effects and outcomes of each option for him/her;

3) Being clear about the relative importance to him/her of the different possible effects and outcomes;

4) Being clear about the chances of the different effects and outcomes happening to him/her, including the uncertainties surrounding the best estimates of them;

5) Being able to trust that the information given to him/her was the best possible;

6) Feeling that he/she has received the level of support and consideration wanted throughout the decision making process, especially in regard to communicating at his level;

7) Feeling that he/she is in control of the decision to the extent he/she wished;

8) Being committed to acting on the decision taken.

To calculate the MDQ score associated with the delivery of a template-based PDA, the subject is first asked to weigh, on a 0-10 scale, the importance to him/her of each of the above eight criteria. The 10 possible levels of importance for each MDQ decision quality criterion, from "not important" to "extremely important" and the related weights are shown in Table 3.6.

Table 3.6. Possible levels of importance and corresponding weights of the decision quality criteria (MDQ)

Levels of importance	**Weights**
Not important (=0)	0
1	0.1
2	0.2
3	0.3
4	0.4
Moderately important (=5)	0.5
6	0.6
7	0.7
8	0.8
9	0.9
Extremely important (=10)	1

Next, the subject is asked to rate how well the use of the PDA has achieved each MDQ criterion, on a scale between 0 and 10, where 0 is “extremely poorly” and 10 is “extremely well” (see Table 3.7).

Table 3.7. Ratings on the decision criteria (MDQ)

How well has the PDA achieved each criterion?	**Ratings**
Extremely poorly (=0)	0
1	0.1
2	0.2
3	0.3
4	0.4
Moderately (=5)	0.5
6	0.6
7	0.7
8	0.8
9	0.9
Extremely well (=10)	1

Multiplying each weight by each rating for each MDQ criterion and adding across criteria, the MDQ score is calculated. The MDQ score ranges between 0 and 1.

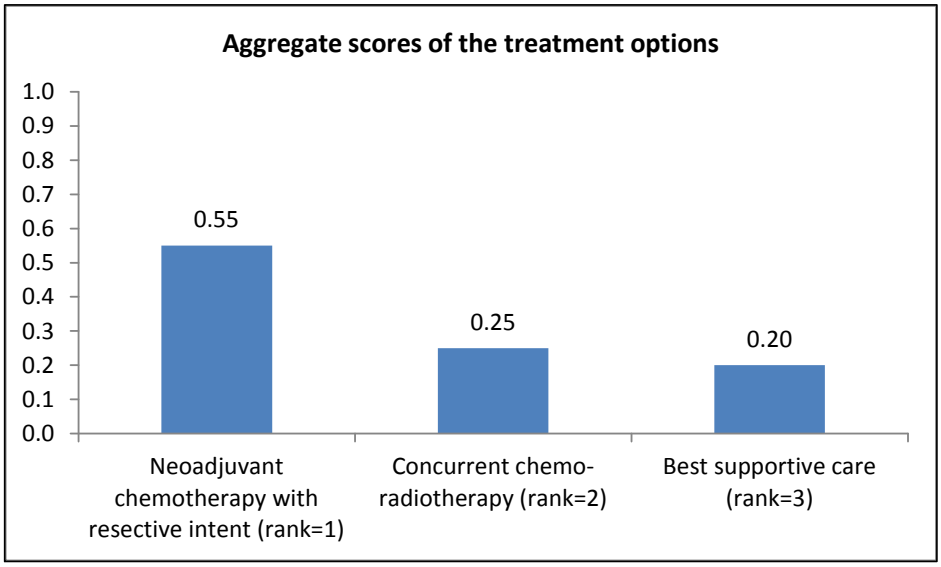
In each hypothetical consultation in hospitals H1 and H2, both the proxy patient and the clinician used the MDQ instrument to assess the quality of the proxy patient's decision using the relevant template-based PDA.

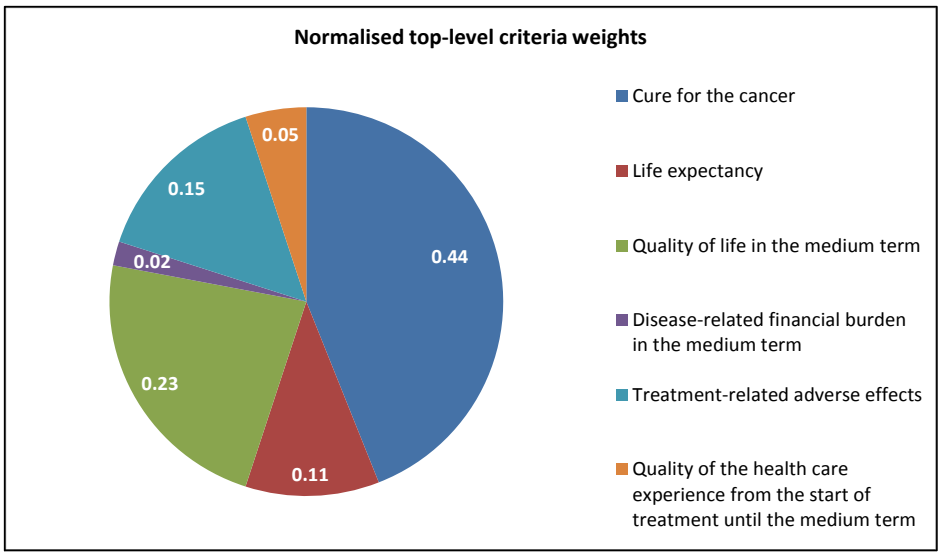
5. Results of delivering the template-based PDAs in the hypothetical consultations.

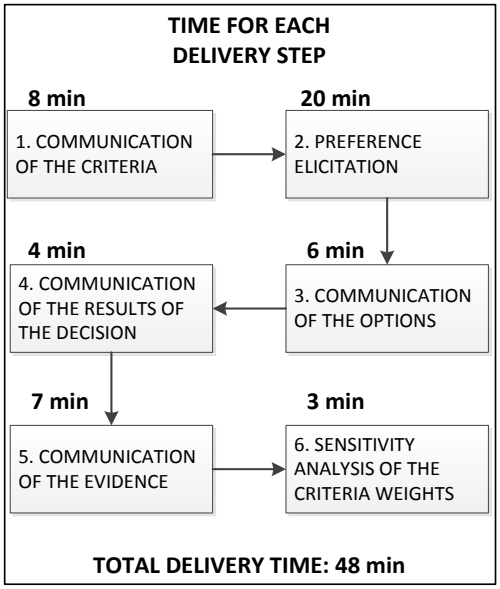
This section summarizes the results of 1) delivering the *Expert Choice*-based PDA in three hypothetical consultations in hospital H1 and of 2) delivering the *ALEL*-based PDAs in three hypothetical consultations in hospital H2. For each hypothetical consultation, the results reported include: 1) The aggregate scores of the three options and their ranking, 2) the weights assigned by the proxy patient to the criteria, 3) the time required to implement each delivery step of the relevant PDA, 4) the results of the sensitivity analysis, 5) the MDQ scores elicited from the clinician and the proxy patient

Figures 3.2-3.4 show the results of delivering the *Expert Choice*-based PDA in the three hypothetical consultations in hospital H1. Figures 3.5-3.7 show the results of delivering the *ALEL*-based PDA in the three hypothetical consultations in hospital H2.

Figure 3.2. Results of delivering the Expert Choice-based PDA in hypothetical consultation 1 with Clinician 1 and a proxy patient (hospital H1)







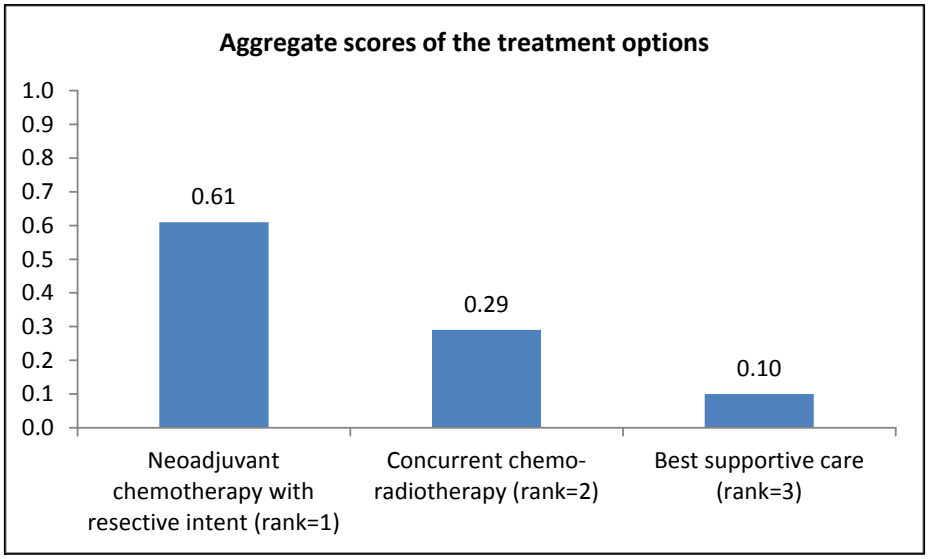
SENSITIVITY ANALYSIS:

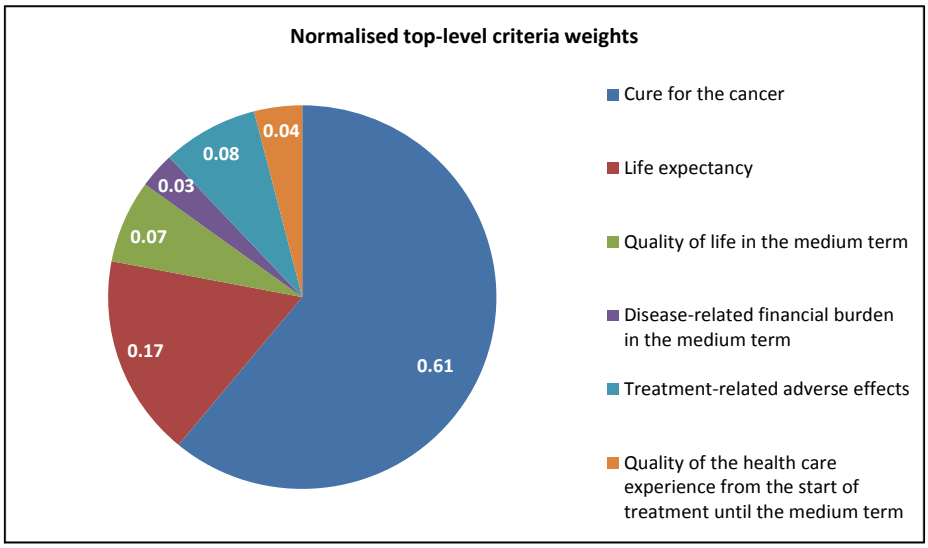
- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

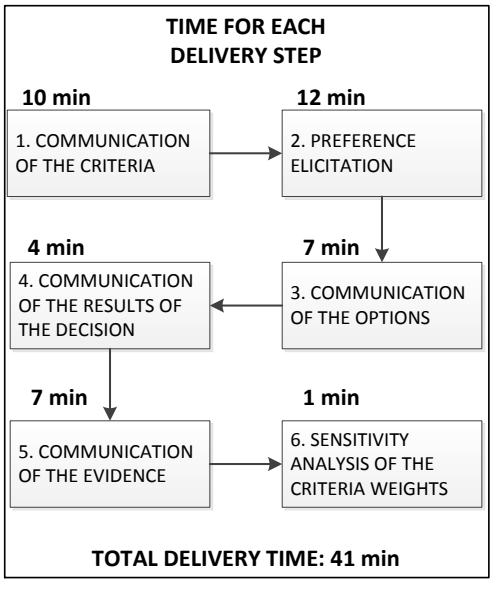
DECISION QUALITY:

- The clinician's MDQ score was **0.67**.
- The proxy patient's MDQ score was **0.75**

Figure 3.3. Results of delivering the Expert Choice-based PDA in hypothetical consultation 2 with Clinician 2 and a proxy patient (hospital H1)







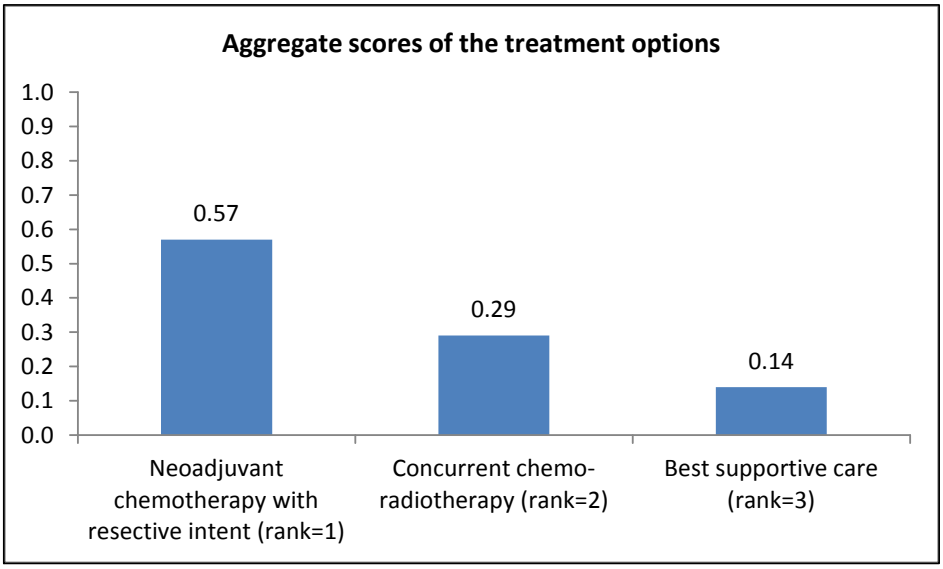
SENSITIVITY ANALYSIS ON THE CRITERIA WEIGHTS:

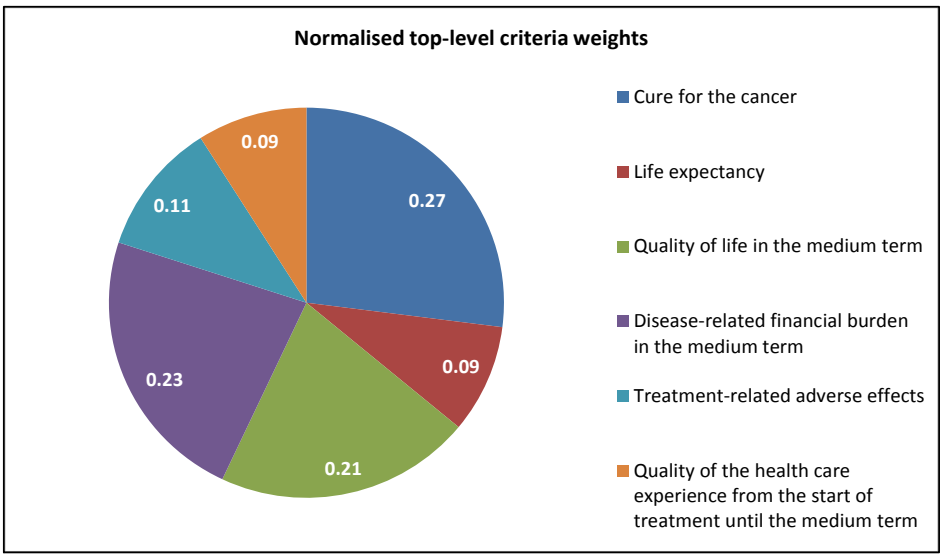
- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

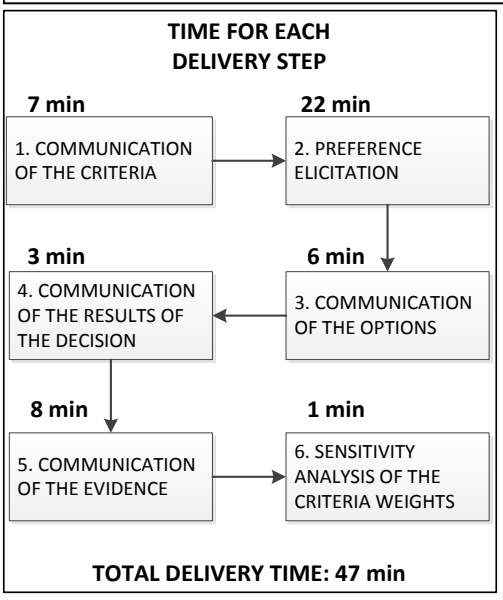
DECISION QUALITY:

- The clinician's MDQ score was **0.77**.
- The proxy patient's MDQ score was **0.70.**

Figure 3.4. Results of delivering the Expert Choice-based PDA in hypothetical consultation 3 with Clinician 3 and a proxy patient (hospital H1)







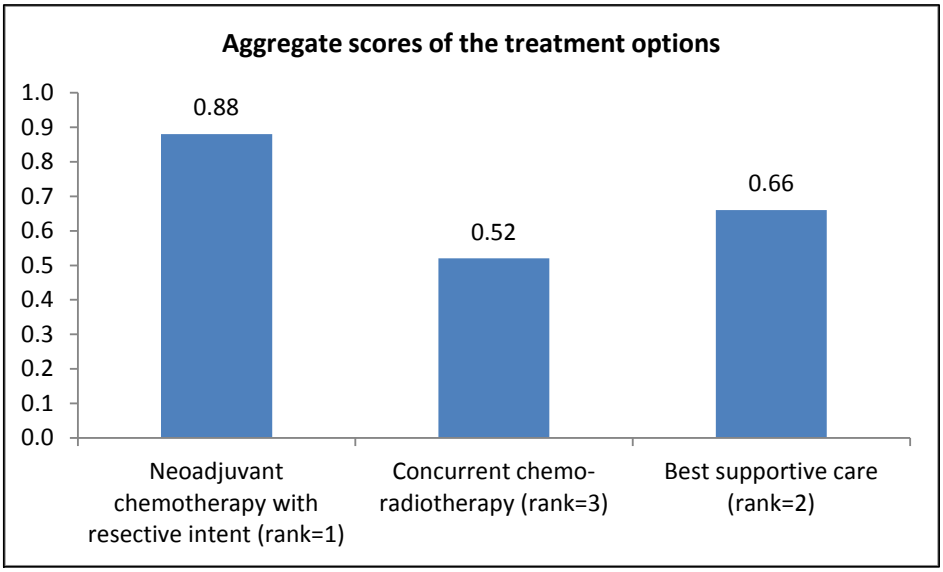
SENSITIVITY ANALYSIS ON THE CRITERIA WEIGHTS:

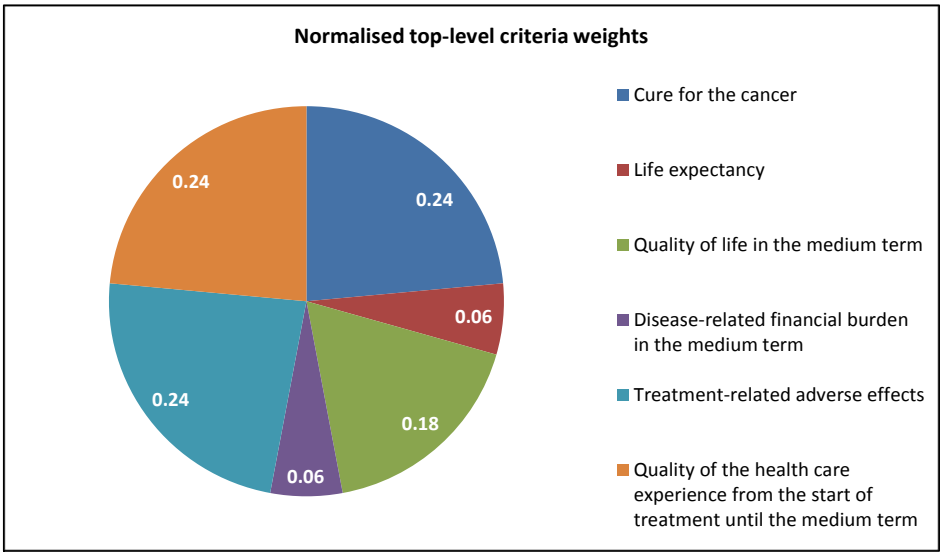
- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

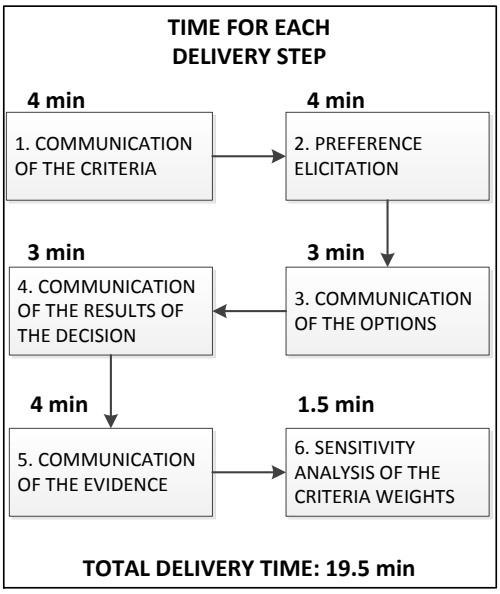
DECISION QUALITY:

- The clinician's MDQ score was **0.60;**
- The proxy patient's MDQ score was **0.78**

Figure 3.5. Results of delivering the ALEL-based PDA in hypothetical consultation 1 with Clinician 1 and a proxy patient (hospital H2)







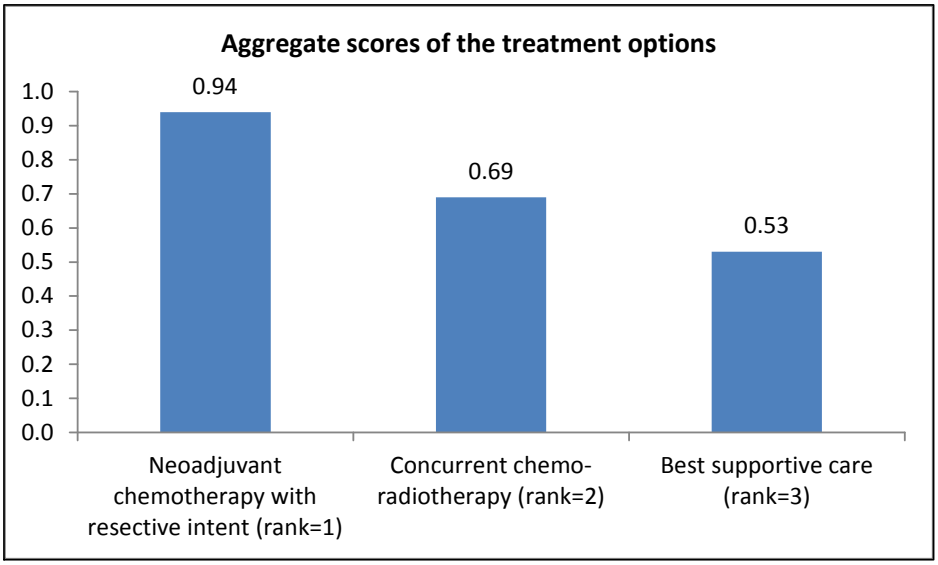
SENSITIVITY ANALYSIS ON THE CRITERIA WEIGHTS:

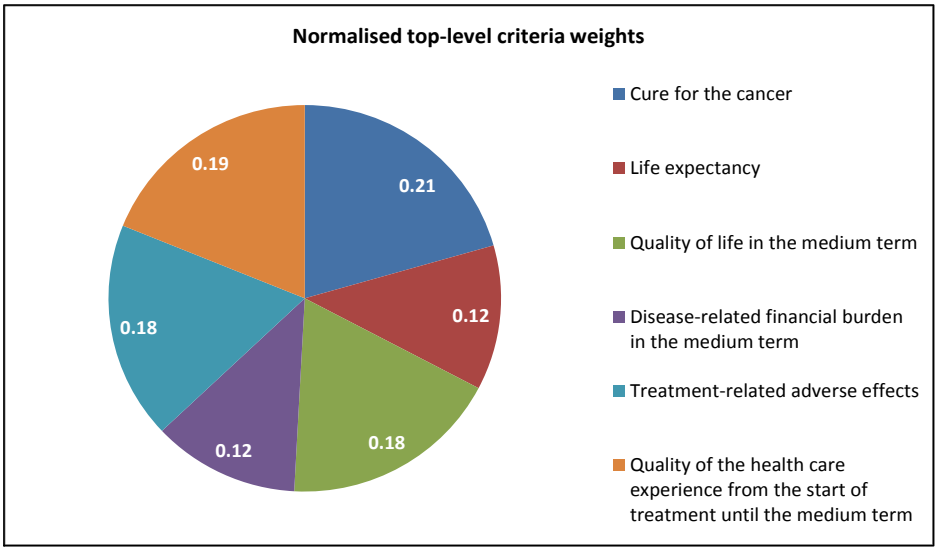
- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

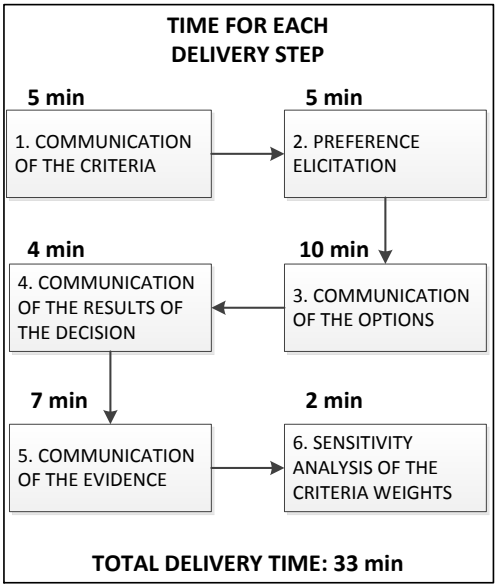
DECISION QUALITY:

- The clinician's MDQ score was **0.48**.
- The proxy patient's MDQ score was **0.61**

Figure 3.6. Results of delivering the ALEL-based PDA in hypothetical consultation 2 with Clinician 2 and a proxy patient (hospital H2)







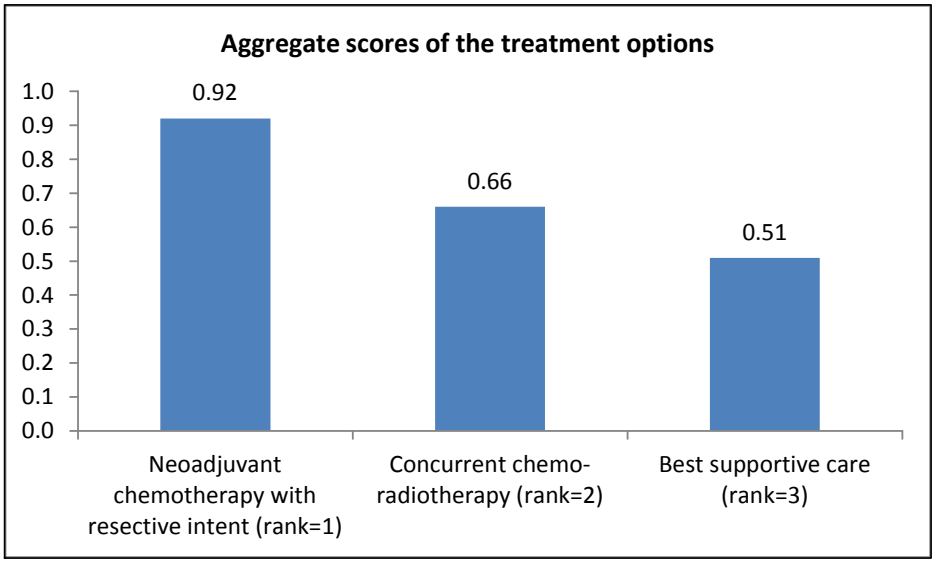
SENSITIVITY ANALYSIS ON THE CRITERIA WEIGHTS:

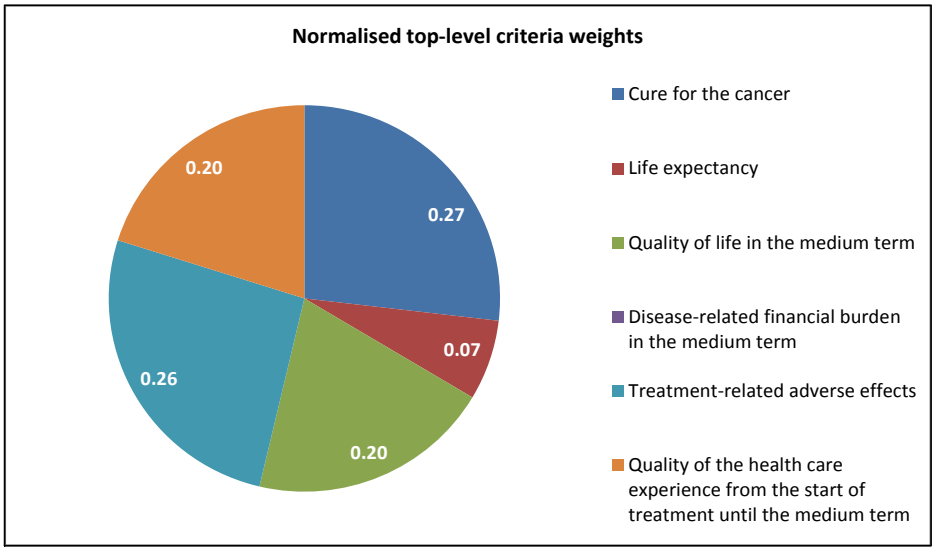
- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

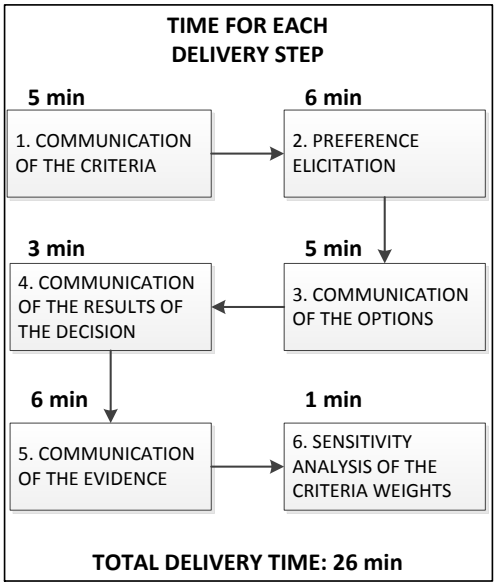
DECISION QUALITY:

- The clinician's MDQ score was **0.88**.
- The proxy patient's MDQ score was **0.87**

Figure 3.7. Results of delivering the ALEL-based PDA in hypothetical consultation 3 with Clinician 3 and a proxy patient (hospital H2)







SENSITIVITY ANALYSIS ON THE CRITERIA WEIGHTS:

- The ranking of the treatment options was not altered by a change of plus/minus 20% in each of the weights

DECISION QUALITY:

- The clinician's MDQ score was **0.75**.
- The proxy patient's MDQ score was **0.81**

6. Analysis.

6.1. Comparing the process of delivering the Stage IIIA$_3$ non-small cell PDA with clinicians in hypothetical consultations in hospital H1 and in hospital H2.

The process of delivering the PDA with clinicians using *Expert Choice* in hospital H1 and using *ALEL* in hospital H2 had both similarities and differences. These similarities and differences are presented for each of the PDA delivery steps: 1) communication of the criteria, 2) preference elicitation, 3) communication of the options, 4) communication of the results, 5) communication of the evidence (if desired by the proxy patient), and 6) sensitivity analysis on criteria weights

One initial, important, difference in the delivery of the PDAs was the level of guidance through each PDA that was required from the clinicians. In the three hypothetical consultations in hospital H1, the clinicians had to switch between *Expert Choice* screens to help the proxy patient through the different PDA delivery steps. This was not the case in the *ALEL*-based PDA. The reason is that the *ALEL*-based PDA was tailored to easily move between screens in a step-by-step manner. Using the Elicia survey functionality, on-screen instructions and information were provided for each delivery step. The proxy patient only had to click on a button to move to the next step. This unique feature in *ALEL* ensured a smooth user experience in the case of the *ALEL*-based PDA.

The communication of the criteria was similar in terms of the criteria described to the proxy patient (the top-level criteria of the hierarchy) and in terms of the language used to describe these criteria. No advantage was taken of the additional functionality available in *Expert Choice* with respect to *ALEL* in terms of being able to present information about all of the criteria in the hierarchy. This is because the three clinicians in hospital H1 considered there would not be enough time to present all the information pertaining to all the criteria during the time available in a standard consultation with a hypothetical lung cancer patient. For the *Expert Choice*-based PDA the time required for the communication of the criteria was, in all hypothetical consultations, longer than for the *ALEL*-based PDA. The reason is that with *Expert Choice* the clinician had to switch between criterion screens (one screen per criterion), while with *ALEL* all the criterion information was presented in the same screen.

The preference elicitation was remarkably different across PDAs. Using the *Expert Choice*-based PDA, each clinician had to first explain to the proxy patient how to do pairwise comparisons of relative importance between criteria using the application's graphical interface. Then, the proxy patient had to perform fifteen pairwise comparisons. In two hypothetical consultations there were inconsistencies in the pairwise comparisons, which required a reassessment of these pairwise comparisons. The advantage of the pairwise comparisons was that consistency between judgments of relative importance was achieved in all cases. The disadvantage of *Expert Choice* is double. First, the process of making fifteen pairwise comparisons of relative importance between criteria is long. Second, the interpretation of the six criteria weights was not straightforward. In the Analytic Hierarchy Process (AHP), the MCDA approach supported by *Expert Choice*, each weight should be interpreted as an average score across options for each criterion, a concept that this researcher found difficult to explain in the hypothetical consultations. The time taken for preference elicitation varied across the hypothetical consultations between a minimum of twelve and a maximum of twenty-two minutes.

Using the *ALEL*-based PDA, the preference elicitation was done in two steps in each hypothetical consultation: first, assigning verbal levels of importance to each of the six top-level criteria using an Elicia screen; then, visualizing these weights as sliding bars on an Annalisa screen to allow for their adjustment. All proxy patients adjusted the criteria weights by small amounts. The interpretation of the weights was that each criterion weight expressed the relative importance of that criterion with respect to the other criteria. The proxy patients had no trouble understanding this definition of weights. During the preference elicitation, the proxy patients made use of the criterion information pop-ups to remember the basic definitions of each of the criteria. The time taken for preference elicitation varied across the hypothetical consultations between a minimum of four and a maximum of six minutes, substantially less than for the *Expert Choice*-based PDA.

The communication of the options was similar for both PDAs in terms of the textual content. The main difference in this stage was that for the *Expert Choice*-based PDA the clinician had to switch across screens to display the information for each option, while in *ALEL* all the information was presented in one Elicia screen. The time taken to complete this stage of PDA delivery varied from a minimum of six minutes to a

maximum of seven minutes for the *Expert Choice*-based PDA and from a minimum of three minutes to a maximum of ten minutes for the *ALEL*-based PDA.

The communication of the results was similar using both PDAs, as both *Expert Choice* and *ALEL* use horizontal bars expressing the overall scores (priorities and ratings, respectively) of the options. *Expert Choice* allows to rank the options on screen, which was found to be a useful feature that all proxy patients took advantage of in the hypothetical consultations in hospital H1. The interpretation of these scores was similar for both *Expert Choice* and *ALEL*. Each option's overall score is the sum of the weighted performances of that option across the six criteria. This was understood clearly by all proxy patients.

All patients chose to visualize the evidence, i.e. the individual scores of the options across the top-level criteria, in all hypothetical consultations. In the *Expert Choice*-based PDA, the priority of each option on a criterion is interpreted as the contribution of the option to the achievement of the criterion, a concept that all proxy patients understood. In *ALEL*, the score or rating of an option on a criterion is the performance of that option relative to the highest performing option, which all proxy patients also understood. The time required for communication of the evidence was between a minimum of seven minutes and a maximum of eight minutes for the *Expert Choice*-based PDA and between four minutes and seven minutes for the *ALEL*-based PDA.

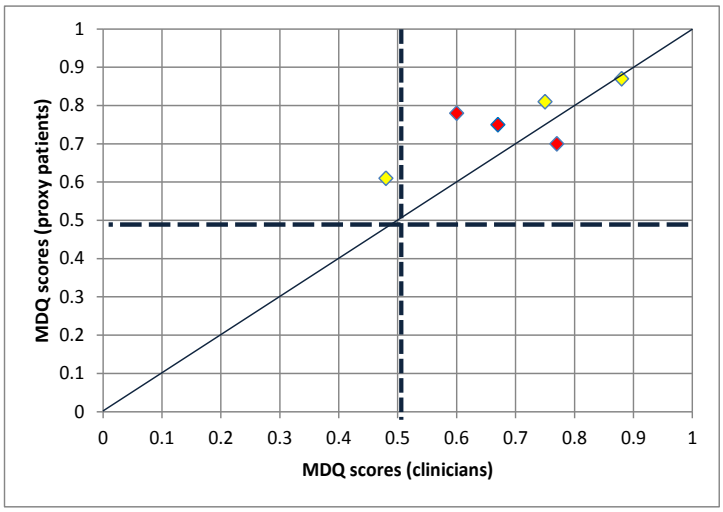
Finally, the sensitivity analysis was performed in a very similar way across hypothetical consultations with the *Expert Choice*-based PDA and the *ALEL*-based PDA. The main difference was that in the hypothetical consultation using *Expert Choice* the clinicians had to switch to a specific sensitivity analysis screen (the Dynamic sensitivity analysis screen), while using *ALEL* the sensitivity analysis was performed directly on the same Annalisa topic screen used for the presentation of the evidence.

6.2. Comparing the quality of the decision across clinicians and proxy patients for the *Expert Choice*-based PDA and the *ALEL*-based PDA.

As mentioned before, after each hypothetical consultation with the *Expert Choice*-based PDA and the *ALEL*-based PDA, both the proxy patient and the clinician were asked to complete the "My Decision Quality" (MDQ) instrument. MDQ produces a

score between 0 and 1 such that a higher score implies higher decision quality. The MDQ score is preference-sensitive, i.e. it depends on the weights assigned by the decision maker to each of the MDQ dimensions. In each hypothetical consultation, the clinician was explicitly asked to complete the MDQ instrument for the proxy patient's decision. In assigning weights for each of the MDQ dimensions, the clinician was asked to consider his own preferences. For example, for the second dimension of the MDQ, "being clear about the possible effects and outcomes of each option for him/her", the clinician was asked to assess how important it was for him/her (i.e. the clinician) that the patient was "clear about the possible effects and outcomes of each option for the patient". Figure 3.8 plots, for each hypothetical consultation, the MDQ score for each proxy patient against the MDQ score of the corresponding clinician. The red points correspond in Figure 3.8 correspond to the MDQ scores for the hypothetical consultations in hospital H1 using the *Expert Choice*-based PDA. The yellow points in Figure 3.8 correspond to the MDQ scores for the hypothetical consultations in hospital H2 using the *ALEL*-based PDA. Points along the diagonal line in Figure 3.8 represent for each hypothetical consultation equal MDQ scores for the clinician and the proxy patient. The dashed lines divide the plot into four quadrants. Points on the top-right quadrant represent levels of decision quality above 0.5 (on a total scale between 0 and 1) for both the proxy patient and the clinician.

Figure 3.8. Plot of the MDQ scores (Expert Choice-based PDA and ALEL-based PDA)



From Figure 3.8, note that in five out of six hypothetical consultations the MDQ scores for both the proxy patients and the clinicians were above the mid-point of the MDQ scale, indicating relatively high levels of perceived decision quality. The highest levels of decision quality corresponded to two of the hypothetical consultations using the *ALEL*-based PDA. From Figure 3.8, note also that four out of the six points lie above the diagonal line. That is, in four out of six hypothetical consultations (two using the *Expert Choice*-based PDA and two using the *ALEL*-based PDA) the proxy patients perceived higher levels of decision quality than did the corresponding clinicians.

6.3. Comparing the overall results in the hypothetical consultations with the *Expert Choice*-based PDA and the *ALEL*-based PDA.

Table 3.8 presents the overall results and rankings of the options on the criteria. From Table 3.8, in five out of six hypothetical consultations neoadjuvant chemotherapy with resective intent (option A_1) was the preferred option, followed by concurrent chemo-radiotherapy (option A_2) and followed by best supportive care. The exception was hypothetical consultation 1 in hospital H2 using the *ALEL*-based PDA. In this consultation, the second most preferred option by the proxy patient was best supportive care.

Table 3.8. Overall scores (ranking of options) for the options in hypothetical consultations

	Hypothetical consultations in Hospital H1			**Hypothetical consultations in hospital H2**		
	1	**2**	**3**	**1**	**2**	**3**
Option A_1	0.55 (1)	0.61 (1)	0.57 (1)	0.88 (1)	0.94 (1)	0.92 (1)
Option A_2	0.25 (2)	0.29 (2)	0.29 (2)	0.52 (3)	0.69 (2)	0.66 (2)
Option A_3	0.20 (3)	0.10 (3)	0.14 (3)	0.66 (2)	0.53 (3)	0.51 (3)

To understand these results it is useful to take a closer look at the highest scoring option on each of the six criteria for each hypothetical consultation and to the

weights assigned to each of these criteria by each proxy patient. This information is shown, respectively, in Table 3.9 and in Figure 3.9.

Table 3.9. Highest scoring option on each criterion for each hypothetical consultations

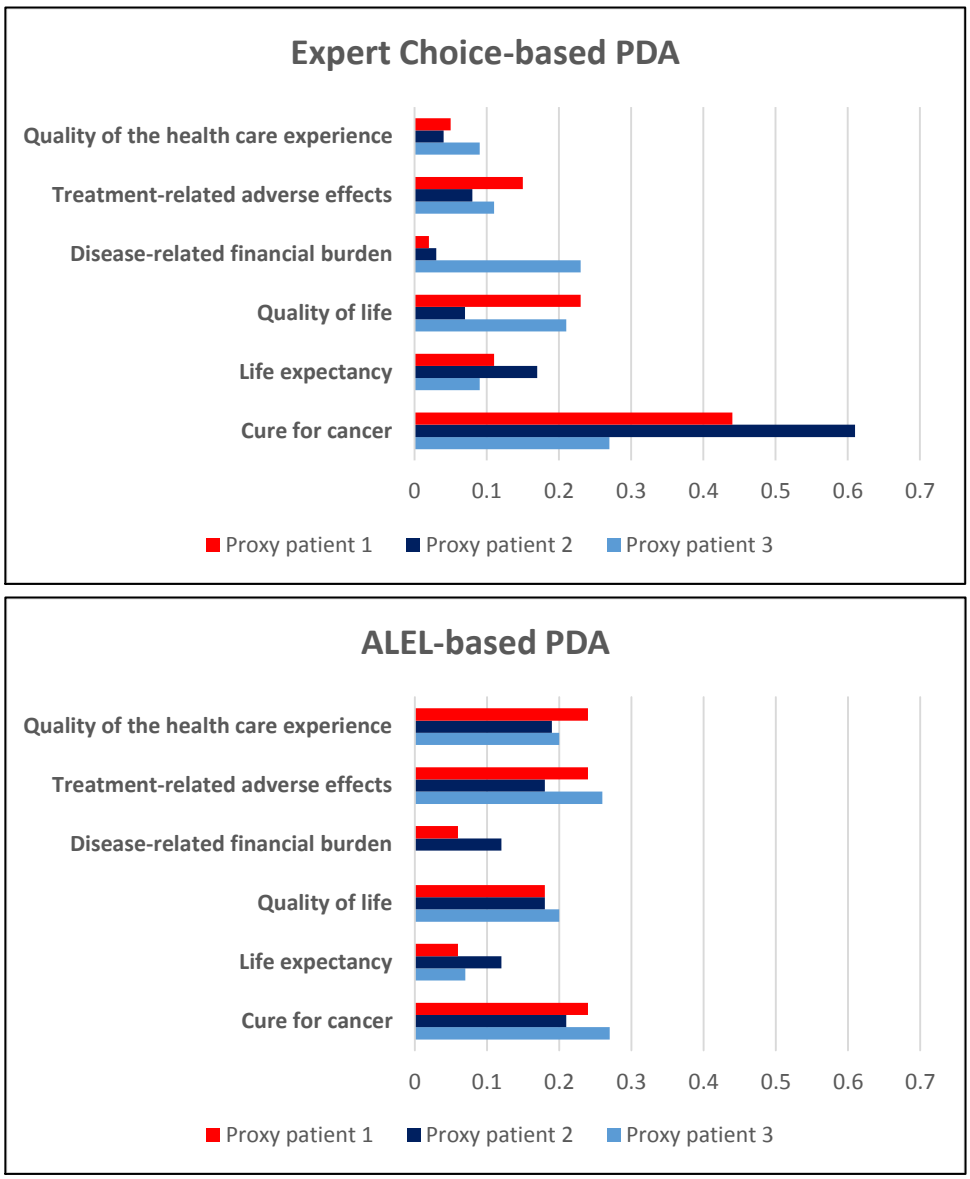
	Hypothetical consultations in Hospital H1 (*Expert Choice*-based PDA)			**Hypothetical consultations in hospital H2 (*ALEL*-based PDA)**		
	1	**2**	**3**	**1**	**2**	**3**
Cure for cancer	A_1	A_1	A_1	A_1	A_1	A_1
Life expectancy	A_1	A_1	A_1	A_1	A_1	A_1
Quality of life in the medium term	A_1	A_1	A_1	A_1	A_1	A_1
Disease-related financial burden in the medium term	A_1	A_1/A_2	A_1	A_1	A_1	A_1/A_3
Treatment-related adverse effects	A_3	A_3	A_3	A_3	A_3	A_3
Quality of the health care experience from start of treatment until the medium term	A_3	A_3	A_3	A_3	A_3	A_3

From Table 3.9, for criteria cure for cancer, life expectancy, and quality of life in the medium term option A_1 is systematically the highest performing option. For criterion

disease-related financial burden in the medium term, this is the case in four out of seven consultations. Conversely, for criteria treatment-related adverse effects and quality of the health care experience from start of treatment until the medium term option A_3 is systematically the highest performing option. Except for disease-related financial burden in the medium term, option A_2 systematically scores between option A_1 and option A_3.

From the above information, the ranking of the options will depend on the relative weights assigned by each proxy patient to the criteria. This information is described in Figure 3.9.

Figure 3.9. Criteria weights assigned by the proxy patients in the hypothetical consultations



From Figure 3.9, top panel, the proxy patients using the *Expert Choice*-based PDA in hospital H1 assigned lower weights to the criteria for which option A_3 was the most preferred, and higher weights to the criteria for which option A_1 was the most preferred. Hence the resulting rankings. From Figure 3.9, bottom panel, the weights assigned to the criteria by the proxy patients using the *ALEL*-based PDA in hospital H2 were more evenly distributed, increasing the likelihood that option A_3 was ranked highest compared to the hypothetical consultations in hospital H1. In particular, proxy patient 1 assigned comparatively high weights to two criteria, quality of the

health care experience between the start of treatment and the medium term and treatment-related adverse effects, for which option A_3 was the most preferred. This led to A_3 ranking higher than A_2 in the case of this proxy patient.

6.4. Exploring the robustness of the overall results in the *Expert Choice*-based PDA and the *ALEL*-based PDA to the independence of irrelevant alternatives.

Recall from Chapter 1 that MCDA approaches such as the AHP and SAW which rely on the normalization of single-criterion scores could lead to rank reversals when options are added to or subtracted from the set of options under consideration. Since both the *Expert Choice*-based PDA and the *ALEL*-based PDA rely on such normalization procedures, it is important to test the impact of adding and subtracting alternatives on the ranking and the overall scores of the options.

Following Belton and Gear [188], an exact copy of option A_2, concurrent chemo-radiotherapy, termed option A_2* was added to each of the models used in the hypothetical consultations. The impact of this addition on the ranking of the options and on their overall scores were explored. In addition, the impact of subsequently eliminating option A_2 and option A_3, best supportive care, on the rankings and overall scores was also explored. The results of these three tests are described in, respectively, Table 3.10, Table 3.11 and Table 3.12.

Table 3.10. Impact on rankings and overall scores of adding option A_2 to the set of options*

	Hypothetical consultations in Hospital H1			**Hypothetical consultations in hospital H2**		
	1	**2**	**3**	**1**	**2**	**3**
Alteration in original ranking	None	None	None	None	None	None
% change in relative overall score of option A_2 with respect to option A_1	-1.2%	-0.3%	+0.0%	+0.0%	+0.0%	+0.0%
% change in relative overall score of option A_3 with respect to option A_1	+2.3%	+7.7%	-2.6%	+0.0%	+0.0%	+0.0%

From Table 3.10, the addition of an exact copy of Option A_2 did not have any undesirable effect on the original rankings of the options obtained in the hypothetical consultations. For the *Expert Choice*-based PDA (implementing the AHP), the relative overall scores of options A_2 and A_3 with respect to option A_1 changed by small amounts. For the *ALEL*-based PDA (implementing SAW with Annalisa), there were no changes in the relative overall scores of options A_2 and A_3 with respect to option A_1

Table 3.11. Impact on rankings and overall scores of subtracting option A_2 from the set of options

	Hypothetical consultations in Hospital H1			**Hypothetical consultations in hospital H2**		
	1	**2**	**3**	**1**	**2**	**3**
Alteration in original ranking	None	None	None	None	None	None
% change in relative overall score of option A_3 with respect to option A_1	+3%	+9%	-2.1%	-2.0%	+0.0%	-5%

From Table 3.11, the deletion of option A_2, concurrent chemo-radiotherapy, did not affect the original rankings in any of the hypothetical consultations. The relative score of option A_3 with respect to A_1 changed over a larger range for the *Expert Choice*-based PDA than for the *ALEL*-based PDA.

Table 3.12. Impact on rankings and overall scores of subtracting option A_3 from the set of options

	Hypothetical consultations in Hospital H1			**Hypothetical consultations in hospital H2**		
	1	**2**	**3**	**1**	**2**	**3**
Alteration in original ranking	None	None	None	None	None	None
% change in relative overall score of option A_2 with respect to option A_1	-3%	-1%	+0%	+6.5%	+1%	+1.7%

From Table 3.12, the deletion of option A_2, concurrent chemo-radiotherapy, did not affect the original rankings in any of the hypothetical consultations. The relative score of option A_3 with respect to A_1 changed over a larger range for the *ALEL*-based PDA than for the *Expert Choice*-based PDA.

Overall, the tests performed show that in this study the results were robust to the addition of an irrelevant alternative and to the subtraction of options. However, there were some changes in the relative scores of the options in the set under consideration when options were added to or subtracted from to the set of options. While this did not lead to illogical results in this particular study, such illogical results could arise in situations where the overall scores of the options are very similar. Note that in the AHP and SAW with Annalisa, the scales on which the overall scores of the options are defined are ratio scales, not interval scales. With ratio scales, differences in overall scores are not meaningful so preference differences of the remaining alternatives were not calculated.

6.5. Implications for clinical practice.

In this study, the delivery of the *Expert Choice*-based PDA and of the *ALEL*-based PDA in clinical consultations was undertaken with proxy patients (non-clinical members of hospital staff) who were healthy individuals. This fact may have biased the choice of clinical management strategy towards the most "aggressive" option, i.e. neoadjuvant chemotherapy with resective intent (option A_1). Option A_1 was the option which performed best out of the three clinical management options in criteria such as cure for cancer, life expectancy or disease-related quality of life. It was also the option which had the strongest treatment-related side effects. The weights assigned by the proxy patients to avoiding adverse effects may have underestimated the weights assigned by actual Stage $IIIA_3$ non-small cell lung cancer patients to that criterion. This possibility helps to highlight that the results of the hypothetical consultations (i.e. the scores and rankings of the options) in hospital H1 and hospital H2 are not directly relevant to clinical practice.

Even though the results of the hypothetical consultations cannot be extrapolated to a real clinical setting, the experience of delivering the *Expert Choice*-based PDA and the *ALEL*-based PDA in hypothetical consultations in hospital H1 and hospital H2 has implications for clinical practice. First, it is feasible to deliver MCDA-based PDAs for the clinical management of Stage $IIIA_3$ non-small cell lung cancer in routine clinical consultations. However, the delivery steps of the PDAs need to be adapted to fit in the time available for consultation with patients. In addition, the amount of information presented in the PDA should be limited to a relatively small number of criteria in order to facilitate understanding. In this study, information on six criteria was well processed by proxy patients.

The introduction of a MCDA-based PDA for the clinical management of lung cancer in routine clinical practice in Spanish NHS hospitals involves a considerable change in the way patients make these decisions with support from their clinicians. Currently, the prevalent mode of decision making is some form of verbal deliberation between the clinician and the patient. With a MCDA-based PDA, all the elements of the decision are made explicit in a quantitative fashion. The patient, with support from the clinician, follows in a step-wise fashion the MCDA process. Patients are not used to make decisions using such a structured, explicit and quantitative decision

technologies. The patients will need to be tutored in the basics of the MCDA approach, with the subsequent implications in terms of time and other resources. A patient recently diagnosed with lung cancer is likely to experience high levels of anxiety, which does not facilitate the tutoring process. Issues of limited health literacy in older patients (recall that they hypothetical patient is a 69 year old individual) will make the tutoring process more difficult.

In the patient-centered mode of shared decision making, the perspective taken in this study, the clinician does not make recommendations, only presents information to the patient using the MCDA-based PDA. Using these tools, clinicians will see their role as agents for the patient diminish, which will not be acceptable to some of them. Patients will see their role as decision makers enhanced, which will not be acceptable to some of them. In general, patients expect a personal interaction with the clinician (this was highlighted several times by the clinicians participating in this study). Some patients will mistrust the use of the MCDA-based PDAs to make such important decisions.

The issues mentioned above place important demands on clinicians. First, they require good clinician-patient communication. Second, they require that clinicians invest time in supporting patients to make decisions using these tools. Considering the workload that clinicians are subject to in routine clinical practice in the Spanish NHS, this is likely to limit the feasibility of implementing MCDA-based PDAs in practice.

MCDA is a resource intensive decision technology. It requires more cognitive effort and time than verbal deliberation to be implemented, particularly for such a complex decision as the clinical management of Stage $IIIA_3$ lung cancer. The limited resources available in routine clinical practice may negatively affect the quality of decision-making if a particular MCDA-based PDA is not delivered ensuring that patients understand well all that is required to help them make the relevant decision using this tool.

The step of preference elicitation is a key part of PDA delivery, as it determines the final ranking of the options. Hence, the impact of imprecise weight estimates on the overall results should be carefully assessed during the PDA delivery process.

Sensitivity analysis plays a very important role in exploring the robustness of the results of the decision to imprecise patient preferences.

Chapter 4: Using a meta-multi-criteria decision model to make the choice between alternative templates for developing and delivering a patient decision aid in routine clinical practice.

Chapter 3 described the methods and results related with Objective 2 of this project: using two MCDA software applications or templates (i.e. *Expert Choice* and *Annalisa in Elicia*) differing in the MCDA approach in order to implement, i.e. deliver, a PDA in an environment replicating actual clinical consultations. This chapter explains the methods and results related with Objective 3 of this research project:

- To build a meta-multi-criteria decision model based on the DRDEA framework and assess the use of this model by clinicians in the Spanish NHS to make the choice between the two MCDA software applications as the basis for a PDA.

The Chapter begins with a justification and an overview of the use of the Decision Resources-Decision Effectiveness Analysis (DRDEA) framework as the basis for choosing the best template for developing and implementing a PDA in routine clinical practice. It then presents an overview of the methods used to develop and implement, with clinicians in the Spanish NHS, a DRDEA-based meta-decision model for the choice between the templates *Expert Choice* and *ALEL* for developing and delivering a Stage $IIIA_3$ non-small cell lung cancer PDA in the context of routine clinical practice. The chapter then describes the process of developing and implementing the DRDEA-based meta-model. It ends with a presentation of the results of the model implementation and the subsequent commentary.

1. Decision Resources- Decision Effectiveness Analysis.

The application of the Decision Resources-Decision Effectiveness Analysis (DRDEA) framework to the choice of MCDA software application or template proposes that the question of choosing one from a set of templates can be framed as a meta-decision problem which is specific to the particular context where the decision is made [19]. The decision problem can be represented by two types of criteria: 1) decision resources (DR) criteria reflecting the resource requirements associated with using each template (e.g. time or cognitive effort required to use the template, as well

as the financial cost associated with implementing the template)[19] and 2) decision effectiveness (DE) criteria expressing the benefits of using the template (e.g. theoretical grounding of the decision-making paradigm underlying the template, or strength and coverage of the evidence included in the template) [19]. In DRDEA, the decision of choosing between alternative templates is preference sensitive and the appropriate analytical technique for making this decision is Multi-Criteria Decision Analysis (MCDA) [19].

The author of this thesis acknowledges that there are other approaches than DRDEA for the task of selecting decision support software applications. For example, those of Le Blanc and Jelassi [231], Ossadnik and Lange [232] and Phillips-Wren et al. [233], all of which have been described in Chapter 1. DRDEA however, is appropriate in the context of this research study because it highlights that the choice between templates for developing and implementing a PDA in routine clinical practice can involve making trade-offs between DR and DE criteria. To illustrate with a basic example, consider the choice, for that task, between a template implementing Multi-Attribute Utility Theory (MAUT) a la Keeney and Raiffa [11] and a template implementing Simple Additive Weighting (SAW). The first template will achieve high levels of DE as it has a strong grounding in axioms of rational decision making, while the second will achieve low levels of DE as it does not have such axiomatic grounding. At the same time, the elicitation of inputs to the decision model in the case of the first template will require a substantial amount of cognitive effort and time, while this will not be the case in the case of the second template. In a context, such as that of routine clinical practice, where time and other resources are scarce, exploring these trade-offs is important.

There are parallels between DRDEA and cost-effectiveness analysis (CEA) [19]. CEA explores the trade-off between cost and effectiveness for the “adoption decision”, i.e. for choosing which health care interventions a health system should adopt. DRDEA explores trade-offs between decision resources and decision effectiveness for the “decision decision”, i.e. deciding how to decide.

2. Overview of methods.

The development and implementation of a DRDEA-based meta-decision model for the choice between *Expert Choice* and *Annalisa in Elicia* (*ALEL*) as the basis for developing and implementing a Stage $IIIA_3$ non-small cell lung cancer PDA followed the following steps:

1. Three clinicians - one pulmonologist (clinician 1), one oncologist (clinician 2), one thoracic surgeon (clinician 3) were recruited from hospital Son Dureta (hospital H3) in Palma de Mallorca;

2. The author of this study, with support from the three clinicians, developed the DRDEA-based meta-decision model using Multi-Attribute Value Theory (MAVT) and the template HiView 3;

3. The DRDEA MAVT model was implemented with the three clinicians from hospital H3.

The DRDEA-based meta-MCDM was developed as proof of concept. The purpose was to test the feasibility of the DRDEA framework in its application to the choice of decision support system.

3. Developing and implementing the meta-decision model.

3.1. Developing the meta-decision model.

As mentioned in the previous section, the meta-decision model was developed using Multi-Attribute Value Theory (MAVT) and the template HiView 3. The decision maker was a hypothetical clinician in charge of developing a PDA, assumed to be the head of a clinical department in a hospital of the Spanish NHS.

The decision question was:

- Which is the best template to develop and deliver a MCDA-based PDA for treating a 69 year old male patient with Stage IIIA3 NSCLC (TNM stage T2N2M0) and cardiovascular and lung comorbidities?

The local decision options under consideration were: *Expert Choice* (option A_1) and *Annalisa in Elicia (ALEL)* (option A_2).

The global decision options included options A_1, A_2 and the following frequently used templates: *Logical Decisions*, *HiView 3, V.I.S.A. Decisions* and *Web-HIPRE*.

The first step in developing the meta-MCDM was defining the relevant hierarchy of decision effectiveness (DE) and decision resources (DR) criteria for choosing between the alternative templates. This step was entirely undertaken by the present author. For this proof-of-concept meta-model, the present author decided that no more than eleven bottom-level DE and DR criteria were to be used in the decision hierarchy of the meta-decision model.

The second step in developing the meta-decision model was calculating the value scores of the two templates on each of the bottom-level criteria of the defined hierarchy. For some of the DE and DR criteria, the scoring of the options was undertaken by the researcher. For some of the DE and DR criteria, this step was undertaken by this researcher in collaboration with each of the three clinicians from hospital H3. Global scales were used to define each of the single-attribute value functions of the bottom-level criteria of the hierarchy. The use of global scales defined each single-attribute value function in tems of the global options considered in the previous section. This ensured that the meta-decision model could be used with decision options other than (or in addition to) *Expert Choice* and *ALEL*.

The following sections describe in detail the resulting meta-decision model.

3.1.1. Determining the relevant decision effectiveness and decision resources criteria

One decision effectiveness criterion and three decision resources criteria were used in the meta-decision model:

1. Normativity in the development of the PDA (DE criterion);

2. Practicality in the delivery of the of the PDA (DR criterion);

3. Ease of use of the template (DR criterion);

4. Acceptability of the PDA (DR criterion).

Each of these four criteria were defined in terms of sub-criteria, as described below.

3.1.1.1. Defining normativity in the development of the MCDA-based PDA.

This DE criterion was defined in terms of a hierarchy of sub-criteria with two levels. Level 1 of the hierarchy was composed of three sub-criteria:

1. Normativity in the structure of the decision model. This sub-criterion is defined as the extent to which the MCDA template is able to incorporate a well-constructed set of criteria in the PDA, i.e. a set of criteria built according to best practice standards. These best practice standards require the criteria to fulfil the properties of completeness, operationality, decomposability, absence of redundancy, and minimum size. For a complex decision like the choice of Stage IIIA3 non-small cell lung cancer treatment, it is very likely that a well-constructed set of criteria will be hierarchical.

2. Normativity in the preference elicitation, defined as the extent to which the template is able to incorporate in the PDA a preference elicitation methodology which reaches normative standards of quality. This criterion was defined in terms of the following two sub-criteria:

2.1. Logical consistency of the preferences, defined as the extent to which the template is able to incorporate in the PDA a preference elicitation technique which produces preferences that are logically consistent.

Logically consistent preferences are generated by preference elicitation techniques which are based on models consistent with rational axioms of decision making. For example, preference elicitation techniques based on MAVT/MAUT. For an in-depth description of these axioms, see Dyer and Keeney and Raiffa;

2.2. Empirical accuracy of the preferences, defined as the extent to which the template is able to incorporate in the PDA a preference elicitation technique which produces preferences that are empirically accurate.

Empirically accurate preferences are obtained by a preference elicitation technique when these preferences correspond to the decision maker's true preferences. However, it is very difficult to objectively know a decision maker's true preferences as they are inherently subjective. In addition, they may not be defined or stable. They may indeed be constructed as the decision maker considers a specific decision situation. Assuming that these true preferences exist, establishing their empirical accuracy is not a straightforward task because it is difficult to find external/objective criteria for validating a given preference elicitation procedure. There are different approaches to testing the experimental validity of multi-criteria preference elicitation procedures using second-best strategies to objective empirical validation, such as convergent, predictive and axiomatic validation. von Winterfeldt and Edwards provide a summary of these approaches as they were used to test the experimental validity of MAUT.

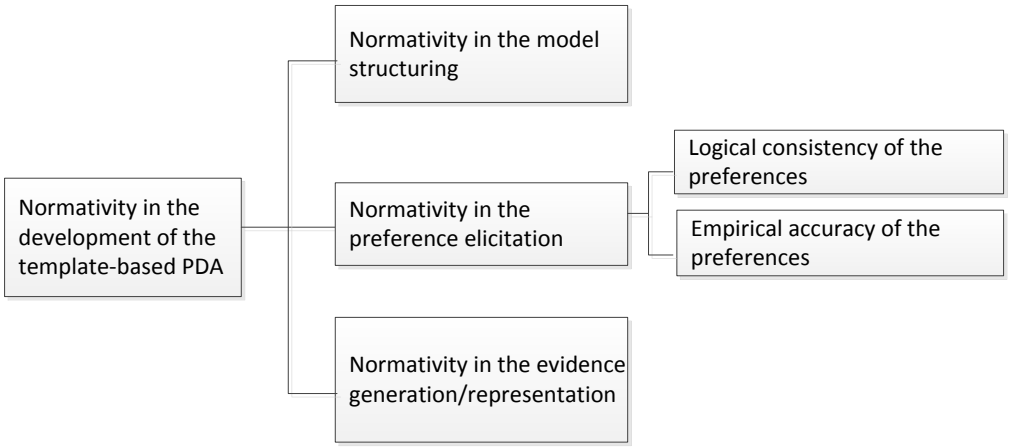
The use of second-best experimental validation approaches such as those mentioned above has shown that MAUT approaches have at least partial experimental validity A similar case is argued by Saaty for the AHP However, these approaches have not set a gold standard for which preference elicitation method yields empirical accuracy of the preferences. In a sense, it remains to the decision maker to determine this gold standard, and to measure against this gold standard the extent to which a given template is able to incorporate into the PDA a preference elicitation procedure that achieves (or is close to) this standard. For the purposes of this thesis, the opinion of this researcher is that MAVT/MAUT, due to a more detailed exploration of the decision maker's preferences compared to AHP, is the best practice standard regarding the empirical accuracy of the preferences.

3. Normativity in the evidence generation/ representation, defined as the extent to which the template is able to represent in the PDA the best available evidence in detail for all the bottom-level decision criteria of a decision hierarchy. The best estimates of the impact of the clinical options on the decision criteria might come, for different criteria, from different sources. Highest quality sources include clinical trials and systematic reviews/meta-analyses. For some (perhaps many) criteria there will be no such high quality (or indeed lower quality) studies that may supply these evidence estimates. In these cases, the best available evidence comes from

methodologically sound approaches to eliciting the relevant estimates from experts. Templates that score highly in this criterion must be able to represent the best evidence of the decision alternatives in detail for all the bottom-level decision criteria of a decision hierarchy. These templates will be hierarchical templates.

Figure 4.1 shows the parent criterion normativity in the development of a template-based PDA and its sub-criteria.

Figure 4.1. Normativity in the development of a MCDA-based PDA: parent criterion and sub-criteria



3.1.1.2. Defining practicality in the delivery of the template-based PDA.

The DR criterion practicality in the delivery of the template-based PDA was defined in terms the following three sub-criteria:

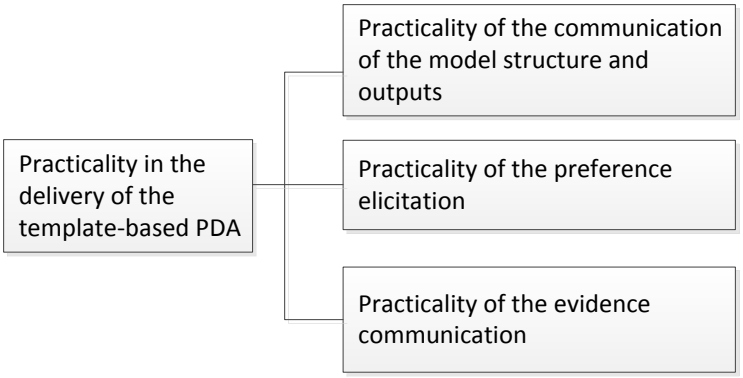
1. Practicality in the communication of the model structure and outputs, defined as the extent to which it is practical, using the template-based PDA, to explain to the patient the model structure (i.e. the set of decision criteria relevant for the clinical decision) and the model outputs (i.e. the aggregate scores of the options) within the constraints of the time available in the consultation.

2. Practicality in the preference elicitation, defined as the extent to which it is practical, using the template-based PDA, to elicit the patient's preferences for the different criteria within the time constraints of a clinical consultation.

3. Practicality in the communication of the evidence, defined as the extent to which the evidence can be explained to the patient using the template-based PDA within the time constraints of a clinical consultation.

Figure 4.2 shows the parent criterion practicality in the delivery of the template-based PDA and its sub-criteria.

Figure 4.2. Practicality in the delivery of the template-based PDA: parent criterion and sub-criteria



3.1.1.3. Defining ease of use of the template interface.

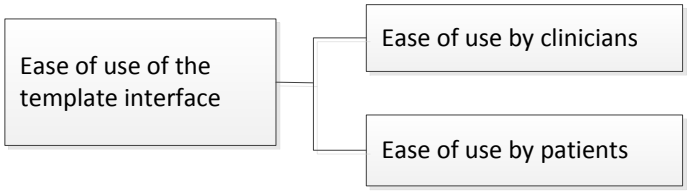
Ease of use of the template interface, a DR criterion, was defined in terms of the following sub-criteria:

1. Ease of use of the template interface by clinicians, defined as the extent to which clinicians will be able to use the interface of a PDA developed using the template with a reasonable and acceptable amount of training;

2. Ease of use of the template interface by patients, similarly defined as the extent to which patients will be able to use the interface of a PDA developed using the template with a reasonable and acceptable amount of training.

Figure 4.3 shows the parent criterion ease of use of the template interface and its sub-criteria.

Figure 4.3. Ease of use of the template interface: parent criterion and sub-criteria



3.1.1.4. Defining acceptability of the template-based PDA.

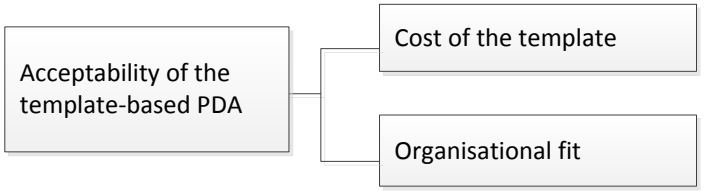
Acceptability of the template-based DA, another DR criterion, is defined in terms of two sub-criteria:

1) Cost of the template, in terms of the cost of a single-user license.

2) Organisational fit of the template-based PDA. This criterion is defined as the extent to which other clinicians in the department perceive that a PDA based on the template will be suited for use in day-to-day clinical practice.

Figure 4.4 shows the parent criterion acceptability of the template-based PDA and its sub-criteria.

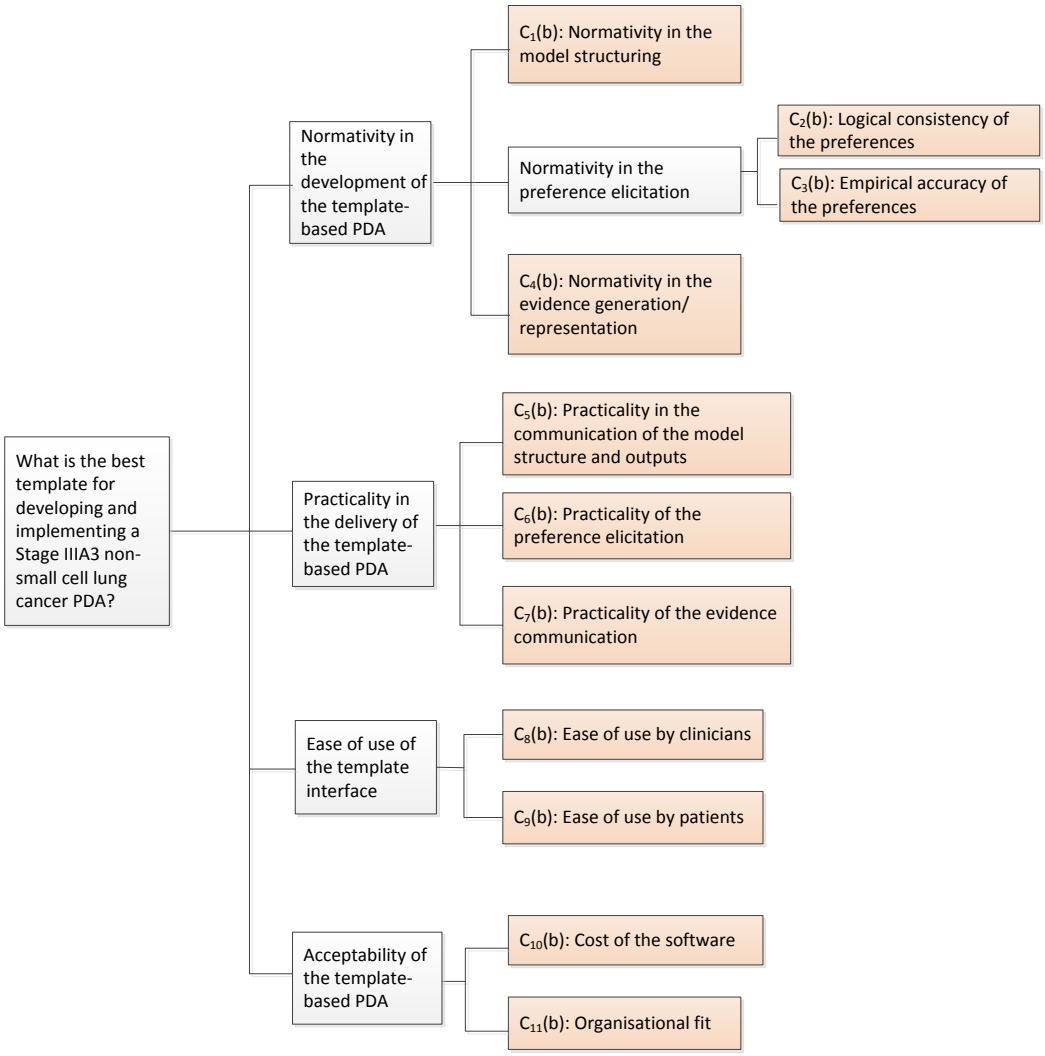
Figure 4.4. Acceptability of the template-based PDA: parent criterion and sub-criteria



3.1.1.5. Full hierarchy of decision criteria for the meta-decision model.

Figure 4.5 shows the final model structure for the meta-decision model. In Figure 4.5, note that the eleven bottom-level criteria are highlighted in red colour and assigned the notation $C_k(b)(k = 1 \dots 11)$.

Figure 4.5. Full hierarchy of decision criteria for the meta-decision model



3.2. Calculating the value scores of the two templates on the bottom-level criteria

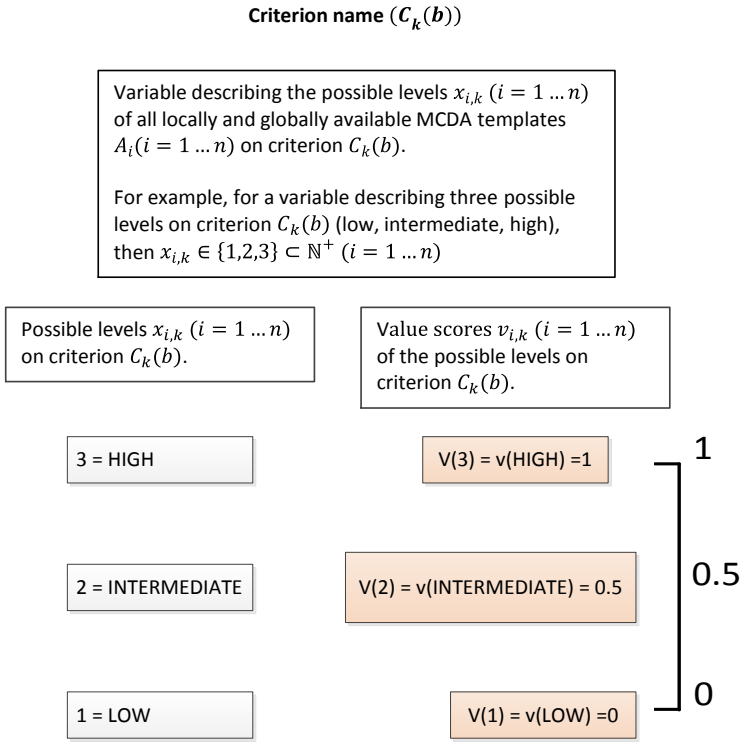
The scores of the two templates A_i $(i = 1,2)$, that is *Expert Choice* and *ALEL*, on the eleven bottom-level criteria $C_k(b)$ $(k = 1 \dots 11)$ were value scores $v_{i,k}$ $(i = 1,2; k = 1 \dots 11)$ on eleven single-criterion value functions. The first stage in calculating these scores for each bottom-level criterion was to build the relevant single-criterion value function. This required the following steps:

1. Defining the variable $x_{i,k}$ $(i = 1 \dots n)$ describing the levels of all templates A_i $(i = 1 \dots n)$ available locally or globally on criterion $C_k(b)$;

2. Assigning value scores $v_{i,k}$ $(i = 1 \dots n)$ to all the levels of any template A_i $(i = 1 \dots n)$ available locally or globally on criterion $C_k(b)$.

For ease of explanation, each single-criterion value function will be illustrated using a diagram. The generic elements of each diagram are illustrated in Figure 4.6.

Figure 4.6. Example of a diagram illustrating each of the single-criterion value functions for criteria $C_k(b)(k = 1 \dots 11)$



As shown in Figure 4.6, at the top of each diagram describing each criterion's single-attribute value function is the variable defining the possible levels on that criterion. Below it on the left is a description of these different levels. Below it on the right is a description (including a 0-1 numerical scale) of the value scores associated with these levels.

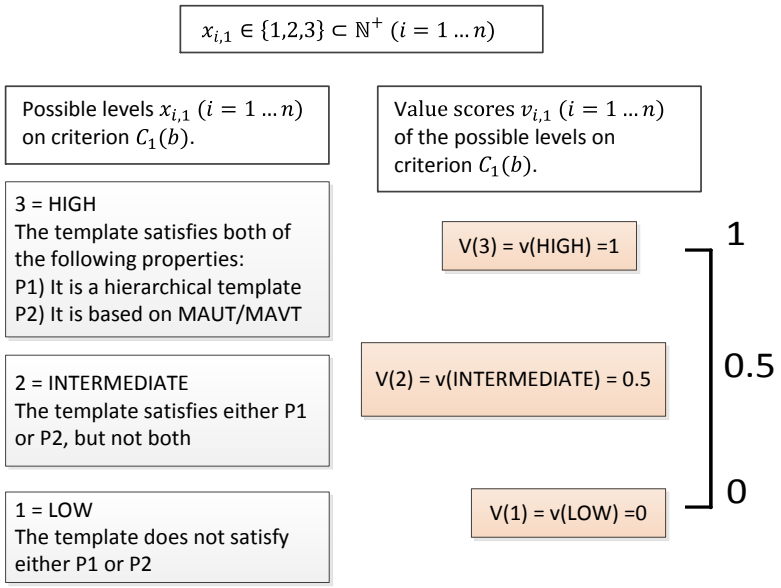
The second step in calculating the value scores of the two MCDA templates $A_i (i = 1,2)$ on each criterion $C_k(b)$ involved calculating the levels $x_{i,k}$ $(i = 1,2)$ of the two options. Mapping these performance levels on the relevant value function led immediately to the value scores $v_{i,k}$ $(i = 1,2)$ of the two options.

The process is now described for each bottom-level criterion.

3.2.1. Normativity in the model structuring $(\boldsymbol{C_1(b)})$.

Normativity in the model structuring was previously defined as the extent to which a template allowed to incorporate in the PDA a well-constructed hierarchy according to the best practice standards. These best practice standards require the set of decision criteria to fulfil the properties of completeness, operationality, decomposability, absence of redundancy, and minimum size. To achieve decomposability, it is important that the template allows for 1) preferences to be expressed for different levels of any particular criterion (as is done in MAUT/MAVT) and for 2) the set of criteria to be expressed as a hierarchy. Figure 4.7 shows the value function for this criterion.

Figure 4.7. Value function for criterion normativity in the model structuring $(C_1(b))$



Looking at Figure 4.7, note that *Expert Choice* is a hierarchical template but not MAUT/MAVT-based, and that *ALEL* is neither. Table 4.1 shows the corresponding levels of the two templates on this criterion and their value scores.

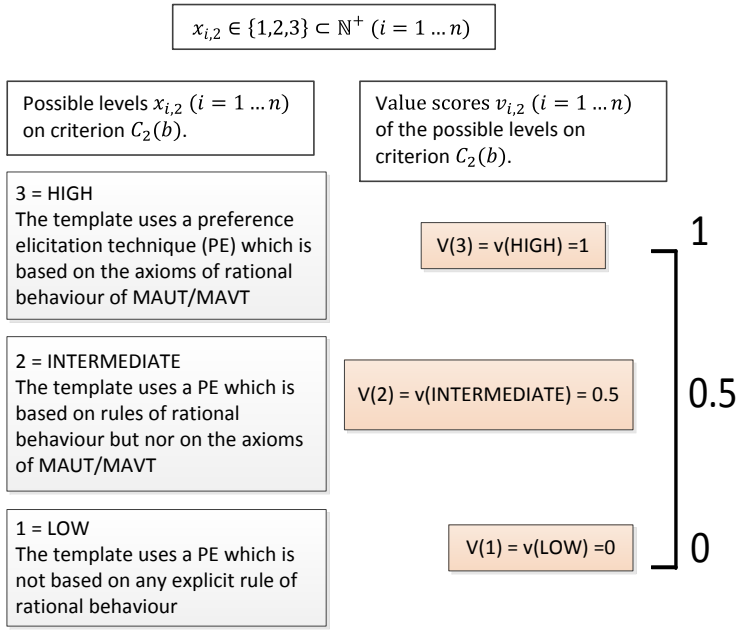
Table 4.1. Levels $x_{i,1}(i = 1,2)$ *and value scores* $v_{i,1}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on criterion normativity in the model structuring* $(C_1(b))$

Option	$\boldsymbol{x_{i,1}}$	$\boldsymbol{v_{i,1}}$
Expert Choice (A_1)	2	0.5
ALEL (A_2)	1	0

3.2.2. Logical consistency of the preferences ($C_2(b)$)

Logical consistency of the preferences was defined as the extent to which a template allowed to incorporate in the PDA a preference elicitation technique which is in accordance with the best standards of decision making. Such standards are the axioms of rational behaviour of MAUT/MAVT. Figure 4.8 shows the value function for this criterion.

Figure 4.8. Value function for criterion logical consistency of the preferences $(C_2(b))$



Expert Choice's preference elicitation technique (pair-wise comparisons of relative importance of the criteria) is based on rules of rational behaviour that ensure the consistency of the patient's preferences (as measured by the consistency index) but not on the axioms of MAUT/MAVT. *ALEL*'s relative importance weight elicitation technique is based on no explicit rules of rational behaviour. Table 4.2 shows the corresponding levels and value scores of the two templates on this criterion.

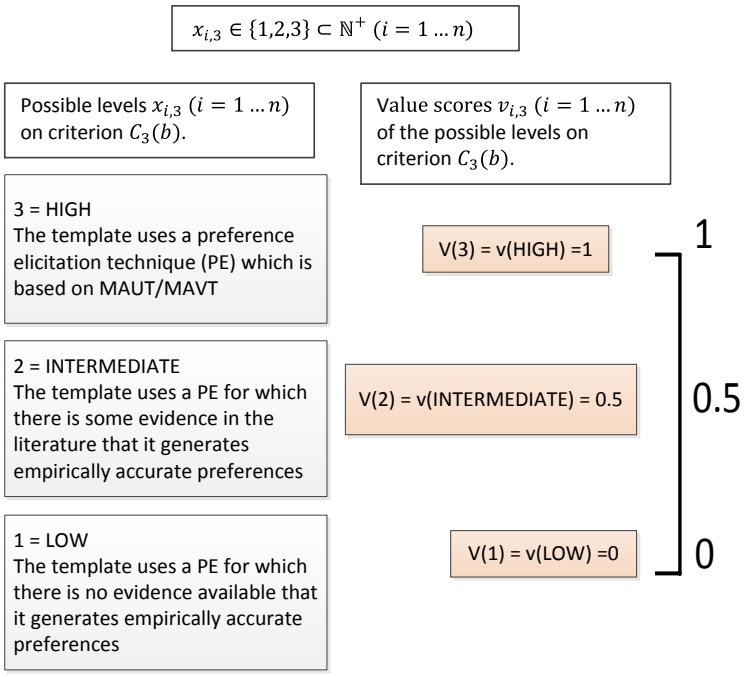
Table 4.2. Levels $x_{i,2}(i = 1,2)$ and value scores $v_{i,2}(i = 1,2)$ of the two templates A_i $(i = 1,2)$ on criterion logical consistency of the preferences $(C_2(b))$

Option	$\boldsymbol{x_{i,2}}$	$\boldsymbol{v_{i,2}}$
Expert Choice (A_1)	2	0.5
ALEL (A_2)	1	0

3.2.3. Empirical accuracy of the preferences $(\boldsymbol{C_3(b)})$.

This criterion was defined as the extent to which a template allows to incorporate in the PDA a preference elicitation technique which is empirically accurate. The empirical accuracy of the preferences is a concept not possible to measure directly. This researcher considered that one first level of empirical accuracy of the preferences was determined by whether or not there was some (indirectly measured) published evidence that a particular preference elicitation technique generated reasonably accurate preferences. This was the case for the preference elicitation techniques of MAUT/MAVT and the AHP. To discriminate whether MAUT/MAVT generated more empirically accurate judgments than AHP (or other approaches) was only possible making a subjective judgment. This researcher considered that MAUT/MAVT preference elicitation techniques generate more empirically accurate preferences because they explore patient preferences more thoroughly than other methods. Figure 4.9 shows the value function for this criterion.

Figure 4.9. Value function for criterion empirical accuracy of the preferences $(C_3(b))$



While there is some evidence in the literature that the preferences generated by the AHP (which is used by *Expert Choice*) are empirically accurate there is not for *ALEL*. Table 4.3 shows the corresponding levels and value scores of the two templates.

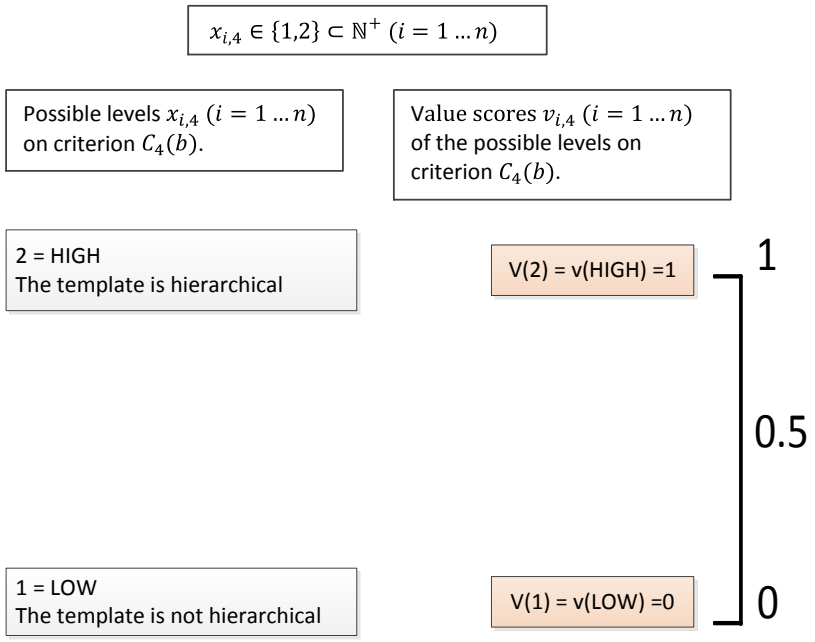
Table 4.3. Levels $x_{i,3}(i = 1,2)$ *and value scores* $v_{i,3}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on empirical accuracy of the preferences* $(C_3(b))$

Option	$x_{i,3}$	$v_{i,3}$
Expert Choice (A_1)	2	0.5
ALEL (A_2)	1	0

3.2.4. Normativity in the evidence generation/ representation $(C_4(b))$

This criterion was defined as the extent to which a template allows for the representation, in the PDA, of the best available evidence on all the bottom-level criteria of the decision model structure. The requirement to achieve this is that the template is hierarchical. Figure 4.10 shows the value function for this criterion.

Figure 4.10. Value function for criterion normativity in the evidence generation/ representation $(C_4(b))$



While *Expert Choice* is a hierarchical template, *ALEL* is not. The corresponding levels and value scores are shown in Table 4.4.

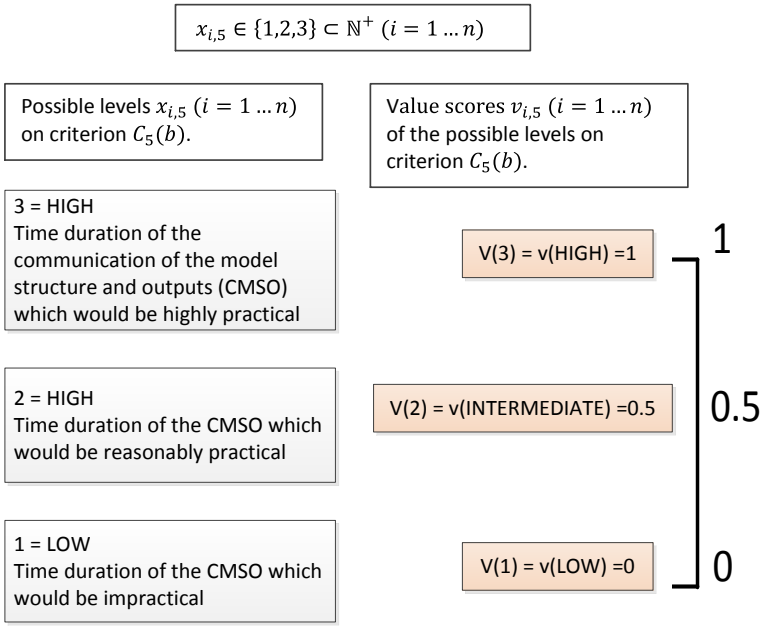
Table 4.4. Levels $x_{i,4}(i = 1,2)$ *and value scores* $v_{i,4}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on normativity in the evidence generation/ representation* $(C_4(b))$

Option	$x_{i,4}$	$v_{i,4}$
Expert Choice (A_1)	2	1
ALEL (A_2)	1	0

3.2.5. Practicality in the communication of the model structure and outputs, or CMSO ($C_5(b)$)

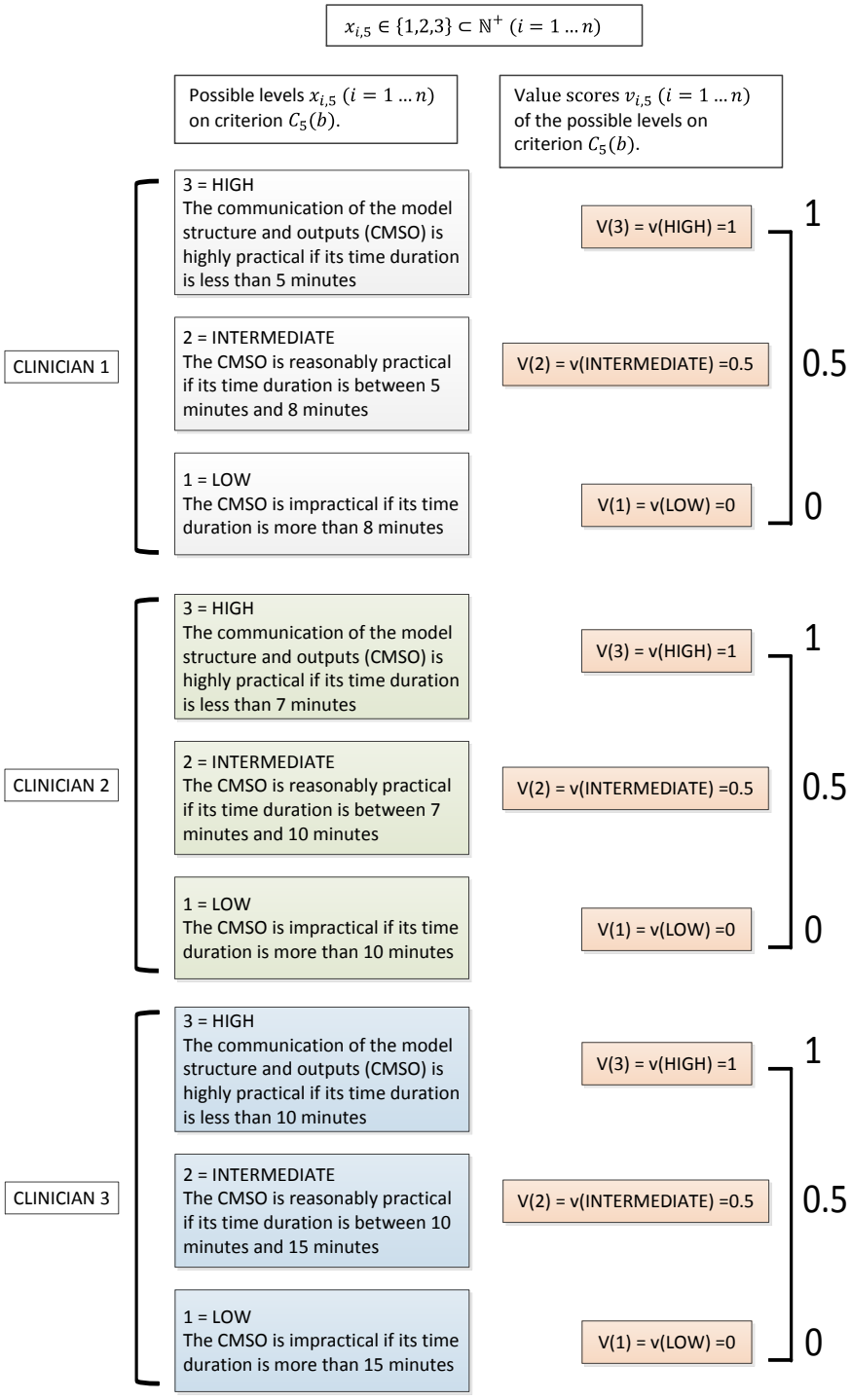
This criterion measured the degree to which it is practical, during the delivery of the PDA, to use a template for explaining to the patient the set of decision criteria and the aggregate scores of the options. This depends on the time it takes to perform this task. Figure 4.11 shows the value function for this criterion.

Figure 4.11. Value function for criterion practicality in the communication of the model structure and outputs ($C_5(b)$)



In Figure 4.11, note that there are no specific time durations assigned to each of the possible levels on this criterion. Each clinician from hospital H3 was asked to supply his/her own estimates of what time duration would make the communication of the model structure and outputs 1) highly practical, 2) reasonably practical and 3) impractical. From each clinician's judgments three value functions with three different time durations were generated for each level (see Figure 4.12).

Figure 4.12. Clinician-specific value functions for criterion practicality in the communication of the model structure and outputs ($C_5(b)$)



As a proxy of the levels of the two templates on this criterion, the following were used: 1) the hypothetical duration of the CMSO using the *Expert Choice*-based PDA if instead of six top-level criteria all of the criteria had been included in the PDA; 2) the duration of the CMSO using the *ALEL*-based PDA.

With respect to 1) the hypothetical duration of the CMSO using the *Expert Choice*-based PDA if instead of six top-level criteria all of the criteria had been included in the PDA, from Chapter 3, the CMSO for the *Expert Choice*-based PDA (with six criteria) took twelve, fourteen, and ten minutes in the hypothetical consultations in hospital H1. These time durations would be at least double if the *Expert Choice*-based PDA included all twenty four criteria. Such time durations would be impractical according to the three clinicians from hospital H3 (see Figure 4.12).

With respect to 2) the duration of the CMSO using the *ALEL*-based PDA, the CMSO for the *ALEL*-based PDA took seven, eight, and nine minutes in each of the three hypothetical consultations in hospital H2. That is, an average of 8 minutes. From Figure 4.11, the level of practicality corresponding to eight minutes was judged by each of the three clinicians in H3 to be, respectively, reasonably practical, reasonably practical and highly practical (see Figure 4.12).

Table 4.5 shows the levels of practicality above and the associated value scores.

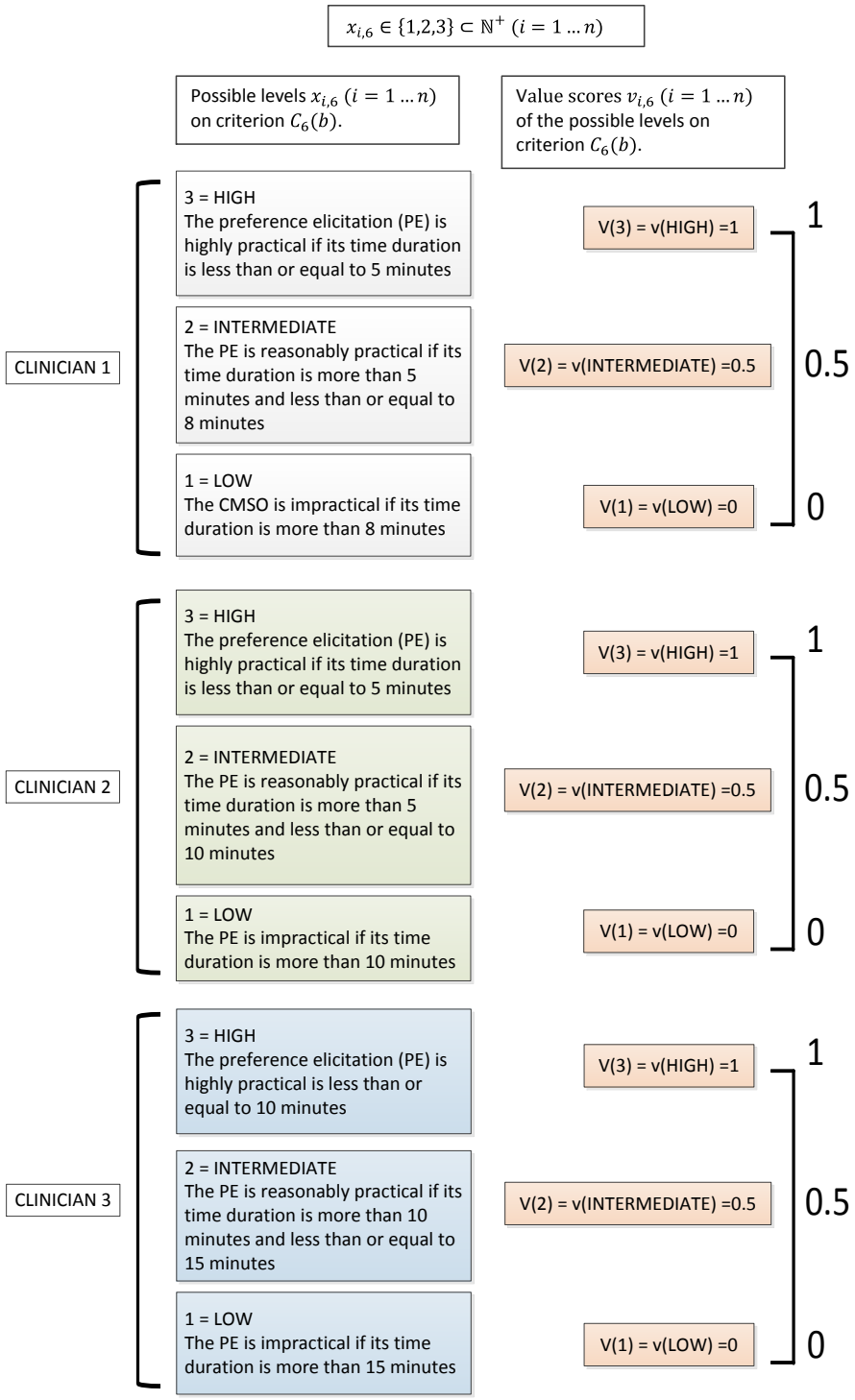
Table 4.5. Levels $x_{i,5}(i = 1,2)$ and value scores $v_{i,5}(i = 1,2)$ of the two templates A_i $(i = 1,2)$ on criterion practicality in the communication of the model structure and outputs $(C_5(b))$

Option	$x_{i,5}$			$v_{i,5}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	1	1	1	0	0	0
ALEL (A_2)	2	2	3	0.5	0.5	1

3.2.6. Practicality in the preference elicitation, or PE $(\boldsymbol{C_6(b)})$

This criterion was defined as the degree to which it is practical to use a template for eliciting the patient's preferences for the different criteria. This depends on the time it takes to perform the preference elicitation task during the delivery of the template-based PDA. Each clinician from hospital H3 was asked to supply his/her own estimates of what time duration would make the preference elicitation 1) highly practical, 2) reasonably practical and 3) impractical. The three resulting value functions are shown in Figure 4.13.

Figure 4.13. Clinician-specific value functions for criterion practicality in the preference elicitation ($(C_6(b))$



As before, 1) the hypothetical duration of the PE using the *Expert Choice*-based PDA if instead of six top-level criteria all of the criteria had been included in the PDA and

2) the duration of the PE using the *ALEL*-based PDA were used as proxies for the level of the two templates on this criterion.

From Chapter 3, the PE for the *Expert Choice*-based PDA (with six criteria) took twenty, twelve, and twenty-two minutes in the hypothetical consultations in hospital H1. These time durations would be much higher if the twenty-four criteria of the Stage IIIA3 NSCLC clinical management decision hierarchy were included in the PDA. According to the three clinicians from hospital H3, such time durations are impractical (see Figure 4.13).

In the hypothetical consultations with the *ALEL*-based PDA in hospital H2, the time duration of the PE was four, five, and six minutes (or five minutes on average). The three clinicians in H3 judged such a time duration to be highly practical.

The levels of practicality above and the associated value scores are shown in Table 4.6.

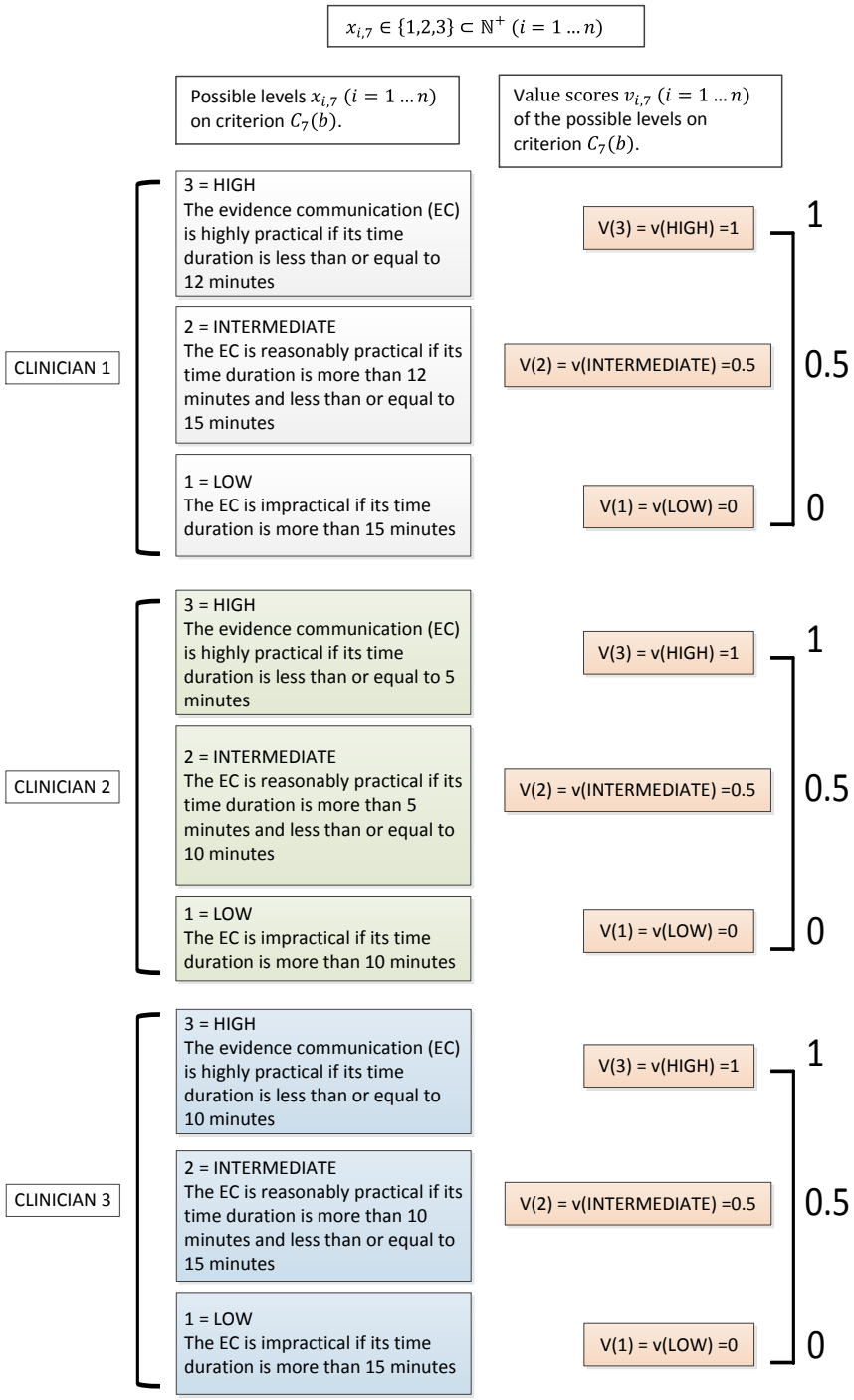
Table 4.6. Levels $x_{i,6}(i = 1,2)$ *and value scores* $v_{i,6}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on criterion practicality in the preference elicitation (*$(C_6(b))$

Option	$x_{i,6}$			$v_{i,6}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	1	1	1	0	0	0
ALEL (A_2)	3	3	3	1	1	1

3.2.7. Practicality in the evidence communication, or EC ($C_7(b)$)

This criterion measured the degree to which it is practical, during the delivery of the PDA, to use a template for explaining to the patient the levels of the options on the criteria. Practicality depends on the time it takes to perform this task during the delivery of the template-based PDA. Each clinician from hospital H3 was asked to supply his/her own estimates of what time duration would make the communication of the evidence 1) highly practical, 2) reasonably practical and 3) impractical. Figure 4.14 shows the three clinician-specific value functions.

Figure 4.14. Clinician-specific value functions for criterion practicality in the preference elicitation ($(C_7(b))$



The levels of the two templates on this criterion were measured using two proxies: 1) the hypothetical time duration of the EC in the delivery of the *Expert Choice*-based

PDA if the evidence had been explained for all the bottom-level criteria and 2) the time duration of the EC in the delivery of the *ALEL*-based PDA.

From Chapter 3, the time duration of the EC in the delivery of the *Expert Choice*-based PDA (with six criteria) was, for each of the three hypothetical consultations in H1, seven, seven and eight minutes. With twenty-four bottom-level criteria it would be about four times as much, a time considered impractical by the three clinicians in hospital H3. The time duration of the EC in the delivery of the *ALEL*-based PDA was, in each of the three hypothetical consultations in hospital H2, four, seven and six minutes, or approximately six minutes on average. This time duration was considered, respectively, highly, reasonably, and highly practical by the three clinicians in hospital H3.

Table 4.7 shows the levels of practicality above and the associated value scores.

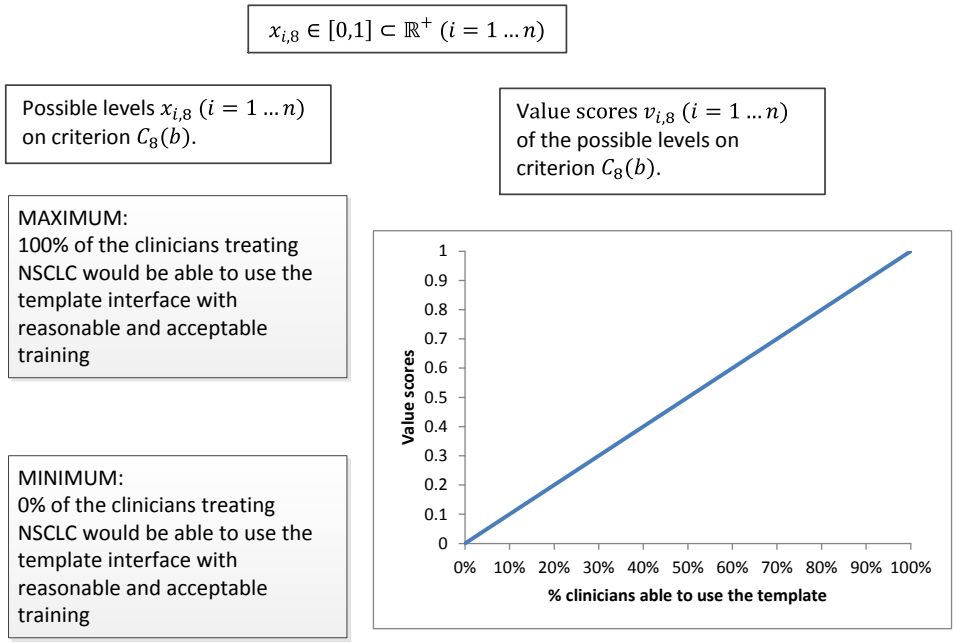
Table 4.7. Levels $x_{i,7}(i = 1,2)$ *and value scores* $v_{i,7}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on criterion practicality in the preference elicitation* $((C_7(b))$

Option	$x_{i,7}$			$v_{i,7}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	1	1	1	0	0	0
ALEL (A_2)	3	2	3	1	0.5	1

3.2.8. Ease of use of the template interface by clinicians ($C_8(b)$)

This criterion was defined as the degree to which, during the delivery of the PDA, a template interface is easy to use by the clinicians. Figure 4.15 shows the value function for this criterion.

Figure 4.15. Value function for criterion ease of use of the template interface by clinicians ($C_8(b)$)



To estimate the levels of the two templates on this criterion, each of the three clinicians in hospital H3 was asked to state the percentage of clinicians treating non-small cell lung cancer (NSCLC) who would be able to use *Expert Choice* and *ALEL* in the delivery of a Stage IIIA3 NSCLC PDA with a reasonable and acceptable amount of training.

The levels of ease of use for clinicians and their associated value scores are shown in Table 4.8.

Table 4.8. Levels $x_{i,8}(i = 1,2)$ and value scores $v_{i,8}(i = 1,2)$ of the two templates A_i $(i = 1,2)$ on criterion ease of use of the template interface by clinicians $(C_8(b))$

Option	$x_{i,8}$			$v_{i,8}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	60%	60%	60%	0.6	0.6	0.6
ALEL (A_2)	80%	70%	75%	0.8	0.7	0.7

3.2.9. Ease of use of the template interface by patients $(\boldsymbol{C_9(b)})$

This criterion was defined as the degree to which, during the delivery of the PDA, a template interface is easy to use by the Stage IIIA3 NSCLC patients. Figure 4.16 shows the value function for this criterion.

Figure 4.16. Value function for criterion ease of use of the template interface by patients $(C_9(b))$

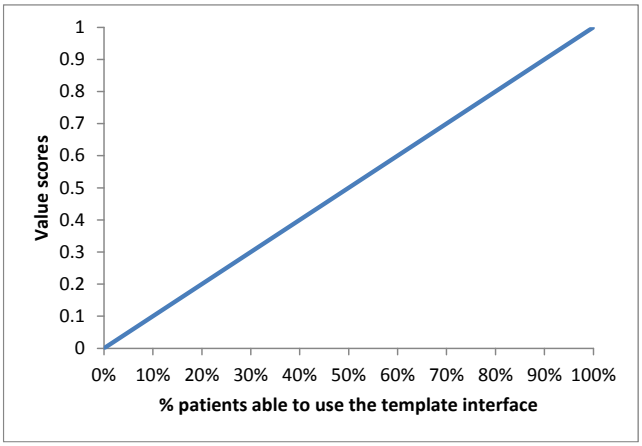
$x_{i,9} \in [0,1] \subset \mathbb{R}^+ \ (i = 1 \dots n)$

Possible levels $x_{i,9}$ $(i = 1 \dots n)$ on criterion $C_9(b)$.

Value scores $v_{i,9}$ $(i = 1 \dots n)$ of the possible levels on criterion $C_9(b)$.

MAXIMUM:
100% of the Stage IIIA3 NSCLC patients would be able to use the template interface with reasonable and acceptable training

MINIMUM:
0% of the Stage IIIA3 NSCLC patients would be able to use the template interface with reasonable and acceptable training



Each of the three clinicians in hospital H3 was asked to state the percentage of Stage IIIA3 NSCLC patients who would be able to use *Expert Choice* and *ALEL* in the delivery of a Stage IIIA3 NSCLC clinical management PDA with a reasonable and acceptable amount of training. Table 4.9 shows the levels of ease of use for patients and their associated value scores.

Table 4.9. Levels $x_{i,9}(i = 1,2)$ and value scores $v_{i,9}(i = 1,2)$ of the two templates A_i $(i = 1,2)$ on criterion ease of use of the template interface by patients $(C_9(b))$

Option	$x_{i,9}$			$v_{i,9}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	20%	20%	40%	0.2	0.2	0.4
ALEL (A_2)	50%	30%	50%	0.5	0.3	0.5

3.2.10. Cost of the template $(\boldsymbol{C_{10}(b)})$.

This criterion was defined as the cost of one user's license of a template. Table 4.10 shows the MCDA templates that were taken into account as global options to develop the value function for this criterion.

Table 4.10. Single user standard licenses available and their cost for six commonly used MCDA templates (in year 2013)

MCDA template	**Type of single-user license**	**Cost (GBP)**
Logical Decisions	Full license for version 7.1	577
HiView 3	Full license for version 3.2.0.7 with 1-year user support	950
V.I.S.A	Full license for V.I.S.A standard version	295
Expert Choice	Full license to *Expert Choice* Desktop	1770
ALEL	1- year managed hosting access to *ALEL*	648
Web-HIPRE	No license required	0

Figure 4.17 shows the value function for this criterion. Note that this value function is linear.

Figure 4.17. Value function for criterion cost of the template $(C_{10}(b))$

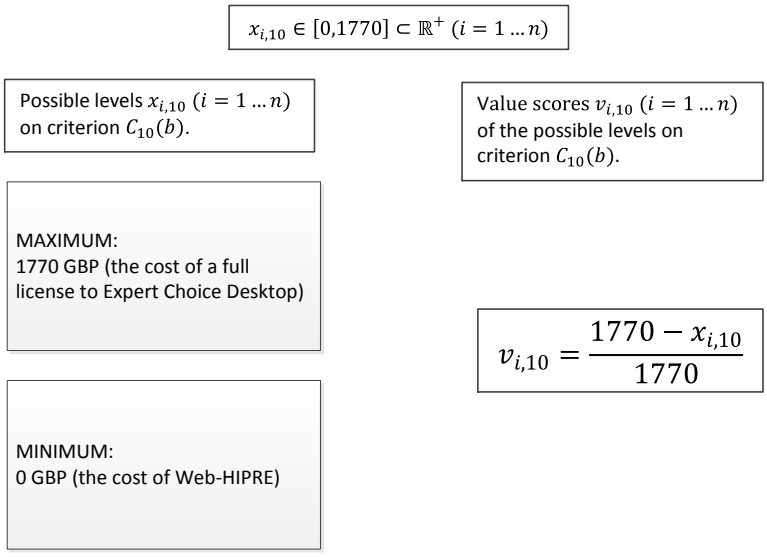


Table 4.11 shows the levels of cost and their associated value scores for the *Expert Choice* and *ALEL* templates.

Table 4.11. Levels $x_{i,10}(i = 1,2)$ *and value scores* $v_{i,10}(i = 1,2)$ *of the two templates* A_i $(i = 1,2)$ *on criterion cost of the template* $(C_{10}(b))$

Option	$x_{i,10}$	$v_{i,10}$
Expert Choice (A_1)	1770	0
ALEL (A_2)	647.88	0.63

3.2.11. Organisational fit $(C_{11}(b))$

This criterion is defined as the extent to which clinicians would find it acceptable to use a template-based Stage $IIIA_3$ NSCLC clinical management PDA in day-to-day clinical practice. To calculate the value scores of the two templates on this criterion, each of the three clinicians in hospital H3 was asked to state the percentage of clinicians involved in treating NSCLC in their hospital who would find it acceptable

to use an *Expert Choice*-based and an *ALEL*-based Stage IIIA3 NSCLC clinical management PDA in day-to-day clinical practice. Figure 4.18 shows the value function for this criterion.

Figure 4.18. Value function for criterion organisational fit $(C_{11}(b))$

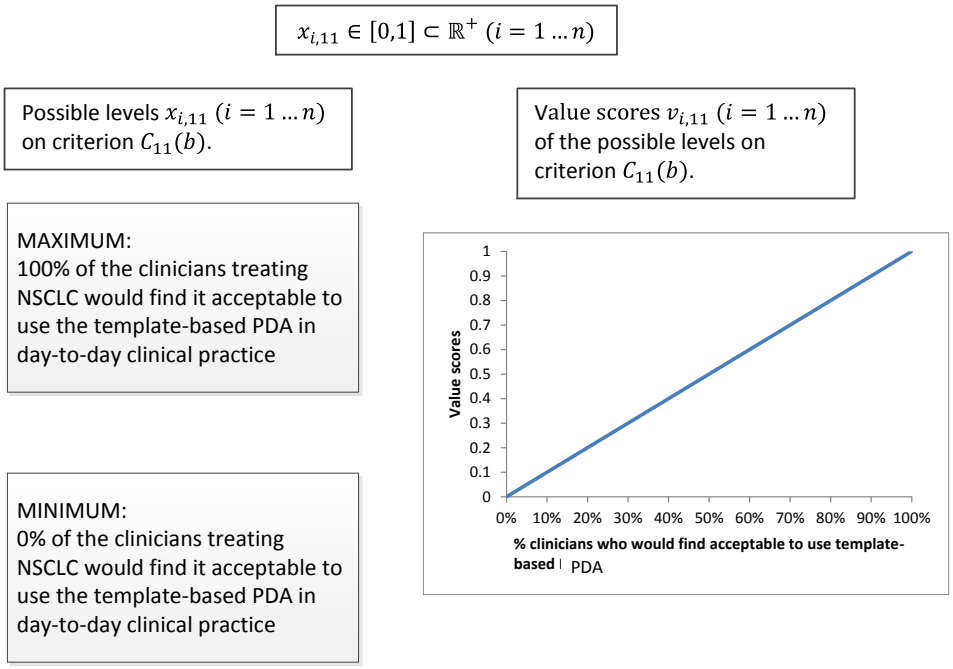


Table 4.12 shows the levels of organisational fit and their associated value scores for the *Expert Choice* and *ALEL* templates.

Table 4.12. Levels $x_{i,11}(i = 1,2)$ and value scores $v_{i,11}(i = 1,2)$ of the two templates A_i $(i = 1,2)$ on criterion organisational fit $(C_{11}(b))$

Option	$x_{i,11}$			$v_{i,11}$		
	Clinician 1	Clinician 2	Clinician 3	Clinician 1	Clinician 2	Clinician 3
Expert Choice (A_1)	25%	10%	50%	0.25	0.10	0.50
ALEL (A_2)	60%	40%	70%	0.60	0.40	0.70

4. Results of solving the meta-MCDM with the three clinicians in hospital H3

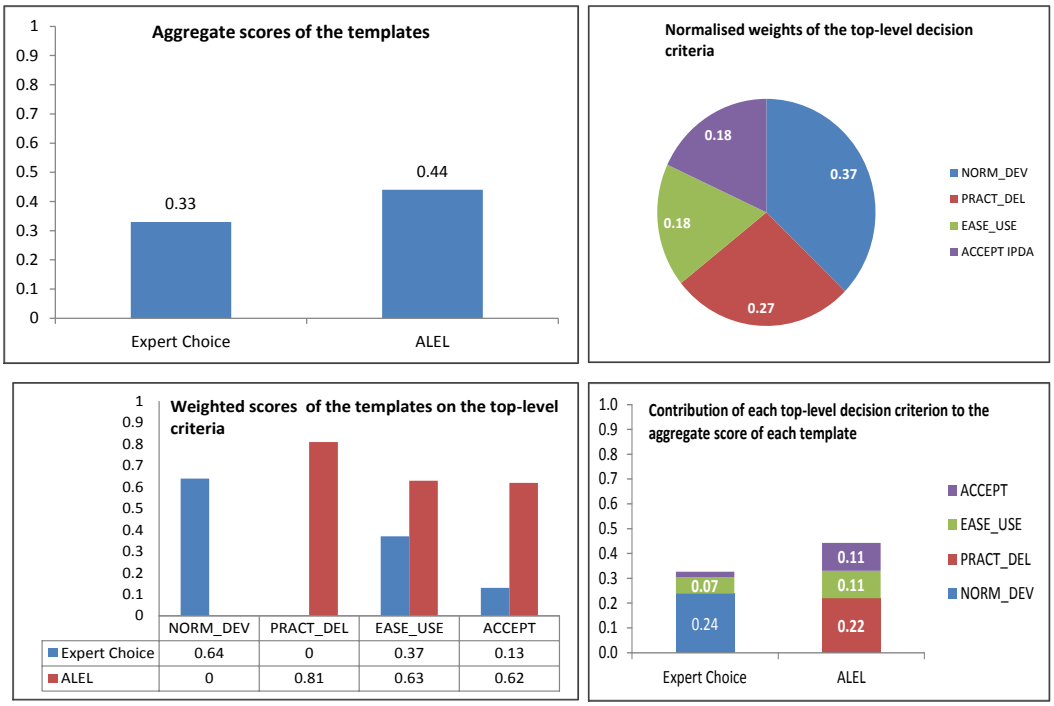
Once the hierarchy of decision effectiveness and decision resource criteria was determined and the scores of the two templates were calculated, both were entered into *HiView 3*. There were three versions of the meta-decision model, one for each of the three clinicians in hospital H3 – the pulmonologist (clinician 1), the oncologist (clinician 2) and the thoracic surgeon (clinician 3). The meta-decison was then solved by each of the three clinicians in an individual meeting with this researcher. Their task was to assess the swing weights of the criteria. Figures 4.18, 4.19 and 4.20 show the results of solving the meta-MCDM for each of the three clinicians.

In the figures, the names of the four top-level criteria of the meta-MCDM hierarchy are shortened in the way shown in Table 4.13.

Table 4.13. Short names of the meta-decision model criteria

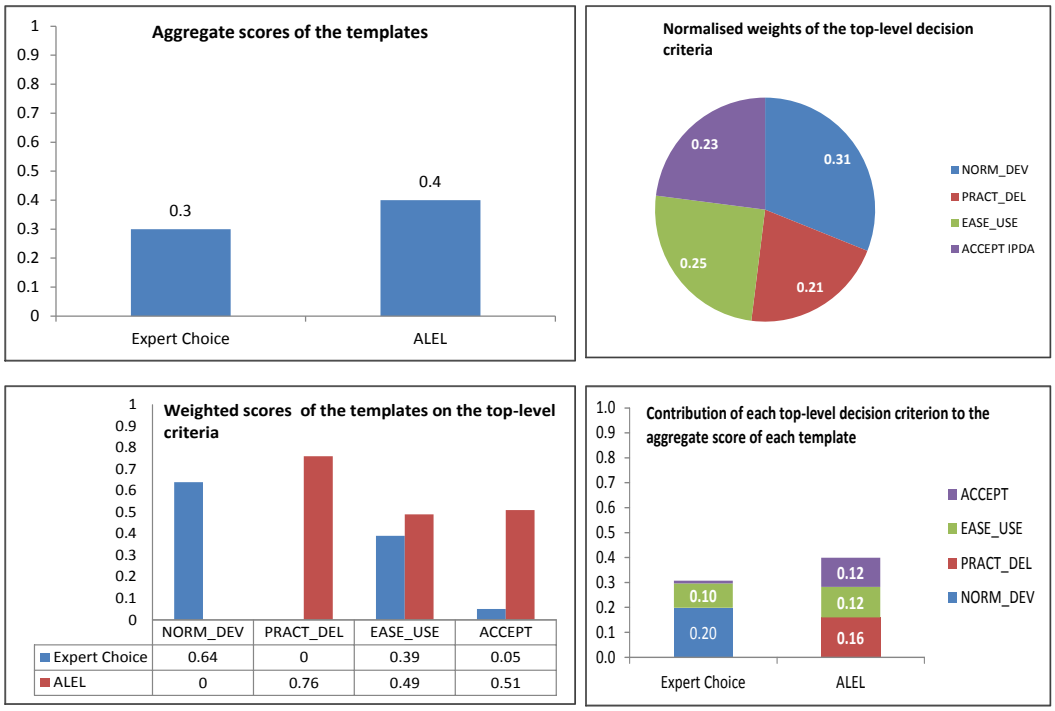
Top-level criteria names	Short name
Normativity in the development of the template-based PDA	NORM_DEV
Practicality in the delivery of the template-based PDA	PRACT_DEL
Ease of use of the template	EASE_USE
Acceptability of the template-based PDA	ACCEPT

Figure 4.19. Results of the meta-decision model (clinician 1)



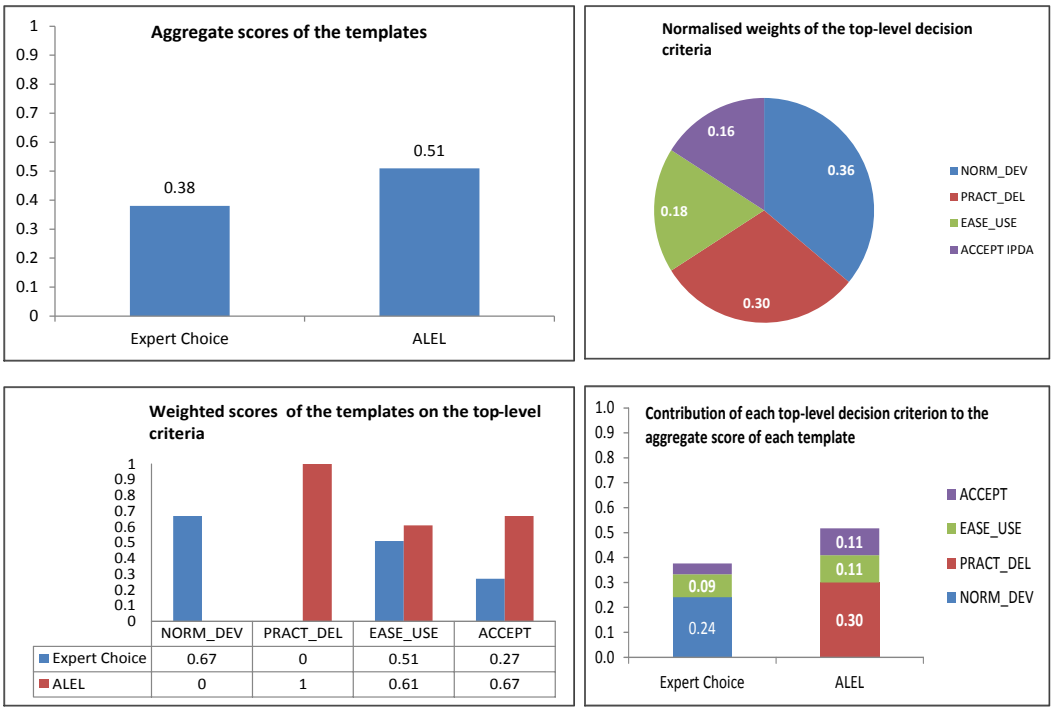
Clinician 1 preferred *ALEL* over *Expert Choice* to develop and deliver a Stage IIIA3 non-small cell lung cancer clinical management PDA, although not by a large margin. The most important criterion for clinician 1 was normativity in the development of the PDA. In Figure 4.19, the bottom left hand panel shows the weighted scores of each of the two templates on the top-level criteria. These weighted scores combine 1) the weights assigned by clinician 1 to the children criteria of these top-level criteria with 2) the scores of the templates on the bottom-level criteria. Note how *Expert Choice* greatly outscores (0.64 to 0) *ALEL* in terms of decision effectiveness (represented by the normativity in the development of the PDA) while *ALEL* greatly outscores *Expert Choice* for two resource criteria (practicality in the delivery of the PDA, acceptability by clinicians of a template based PDA in day-to-day clinical practice) and somewhat less in the third decision resource criterion (ease of use in the template). The bottom-right hand panel in Figure 4.19 shows the contribution of each top-level criterion to the final score of the template. In switching from one template to another, there is a trade-off between normativity in the development of the PDA and practicality in the delivery of the PDA. Switching from *ALEL* to *Expert Choice* involves forfeiting practicality in the delivery of the PDA in exchange for higher normativity. The results were robust to a sensitivity analysis on the weights of all the bottom-level decision criteria.

Figure 4.20. Results of the meta-decision model (clinician 2)



The results of the meta-decision model using clinician 2's preferences and value functions (where relevant) were quite similar to the results using clinician 1's preferences and value functions (where relevant). From Figure 4.20, top left hand panel, *ALEL* was preferred over *Expert Choice*, although not by a large margin. Again normativity in the development of the PDA was the most important criterion in the choice of template. Again a trade-off was evidenced between the decision effectiveness criterion and the decision resource criteria when switching across templates. Perhaps the main difference was that clinician 2 did not consider that the two templates were similar in terms of ease of use in the delivery of the template-based PDA. The results of the meta-decision model for clinician 2 were robust to changes in the weights of the bottom-level criteria.

Figure 4.21. Results of the meta-decision model (clinician 3)



The pattern that was seen for clinicians 1 and 2 was repeated for clinician 3. *ALEL* was preferred over *Expert Choice*, although in this case by a larger margin than clinician 1 and clinician 2. Interestingly, although clinician 3 attached the highest weight of the three clinicians to criterion normativity in the development of the template-based PDA, he also attached the highest weight of the three to criterion practicality in the delivery of the template-based PDA. In the bottom right hand panel of Figure 4.21 note how much criterion practicality in the delivery of the PDA contributes to the score of the *ALEL* template. The results of the meta-decision model for clinician 3 were robust to changes in the weights of the bottom-level criteria.

5. Implications for clinical practice.

From the above results, the MCDA formulation of the Decision Resources-Decision Effectiveness Analysis (DRDEA) framework can be useful to help clinicians choose between alternative templates to develop and deliver a Stage $IIIA_3$ non-small cell lung cancer PDA in the context of day-to-day clinical practice in the Spanish NHS. Specifically, the results of developing and implementing the DRDEA-based meta-decision model showed with three clinicians from hospital H3 show that the choice

of template to develop and implement a PDA indeed involves making trade-offs between decision effectiveness and decision resources. This is not a surprising result considering the time and other resource constraints that clinicians face in the workplace.

The three clinicians with which the meta-decision model was developed and implemented found the use of MAVT with swing weights via the template HiView 3 straightforward and informative. There is no reason to assume that this approach cannot be used by other clinicians for choosing the most appropriate template as the basis for a PDA. MAVT is a MCDA methodology with solid axiomatic grounding in rational decision making. It is recommended here a "best practice" approach for the application of DRDEA to the choice of template as the basis of a PDA.

The three clinicians from hospital H3 considered that the most important criterion for choosing between templates was the normativity in the development of the MCDA-based PDA. Assuming that other clinicians express similar preferences to the clinicians in hospital H3, it is likely that, if resource constraints were not an issue, clinicians would choose templates which implement MCDA approaches that allow for high level of normativity in the development of a MCDA-based PDA. As defined in the proof of concept meta-MCDA model, best practice standards of normativity in the development of a MCDA-based PDA correspond to templates which implement MAVT/MAUT at the same time that they allow for a hierarchical representation of the decision problem. Templates of this type include HiView 3, Logical Decisions, and V.I.S.A Decisions.

However, decision resources as defined in the meta-decision model may impose constraints to the desired levels of decision effectiveness in routine clinical practice. This may be particularly relevant in the case of complex decisions such as the clinical management of cancer. For the particular example researched in this thesis, the clinical management of Stage $IIIA_3$ non-small cell lung cancer, the large number of criteria considered relevant to the decision by clinicians in hospitals H1 and H2 was one of the main sources (although not the only one) of decision resource constraints. This was evidenced by some of the opinions expressed by the clinicians in hospital H3. For example, in terms of practicality in the delivery of the lung cancer PDA, the three clinicians in hospital H3 considered that it was impractical to

communicate to the patient the twenty-four bottom-level criteria of the hierarchy using the *Expert Choice*-based PDA during the time available in clinical consultations. In contrast, two of the clinicians considered that it was reasonably practical to communicate the six top-level criteria of the hierarchy using the *ALEL*-based PDA, and one considered that it was highly practical. With respect to the practicality in the preference elicitation, the three clinicians considered that it was impractical for the patient to perform all the required pairwise comparisons for the full hierarchy of decision criteria for the *Expert Choice*-based PDA during the time available in a clinical consultation. They considered, however, that it was highly practical to elicit preferences over the six top-level criteria of the hierarchy using the approach described in the *ALEL*-based PDA (i.e. first assigning verbal levels of importance to the criteria and then adjusting the resulting criteria using sliding bars). From these comments it seems clear that developing a hierarchy of criteria of minimum size which is still relevant to the decision problem might alleviate some of the decision resources constraints imposed on the development and delivery of MCDA-based PDAs in routine clinical practice.

With respect to the ease of use (or lack thereof) of the templates, the three clinicians considered that 1) a substantial percentage of clinicians treating non-small cell lung cancer (40%) would not find the *Expert Choice* interface easy to use even with training and that 2) between 20% and 30% of clinicians treating non-small cell lung cancer would not find the *ALEL* interface easy to use even with training. These percentages increased for the case of patients. The three clinicians considered that 1) between 60% and 80% of Stage $IIIA_3$ non-small cell lung cancer patients would not find the *Expert Choice* interface easy to use even with training and that 2) between 30% and 50% of these patients would not find the *ALEL* interface easy to use even with training. This may indicate that there is more to the use of a MCDA-based PDA than adequate and acceptable training in its use. Training in the basic concepts of MCDA might also be important.

Finally, with respect to the acceptability of the templates for use by clinicians in their consultations, the three clinicians considered that between 50% and 90% of clinicians treating non-small cell lung cancer would not find it acceptable to use the *Expert Choice* PDA in their day-to-day clinical practice. The same three clinicians

considered that between 30% and 60% of clinicians treating non-small cell lung cancer would not find it acceptable to use the *ALEL* template in their day-to-day clinical practice. Unfortunately, the implications of these statements could not be explored with the three clinicians due to lack of time to participate in this project beyond the implementation of the meta-decision model. From the literature, there are a number of barriers to the implementation of PDAs in clinical practice which could play a part in the acceptability of these tools. Clinicians cite scepticism about the content of PDAs [105], lack of training in shared decision-making and in the use of PDAs [7], concerns about the adequacy of PDAs for some groups of patients (e.g. vulnerable patients such as the elderly or patients with little education) [7], and competing demands and time constraints [105].

Chapter 5: Discussion

1. Main findings

Undertaken in close collaboration with clinicians in the Spanish NHS, the research undertaken in this study established that MCDA is a potentially useful technique for the development and implementation, i.e. delivery, using dissimilar templates, of MCDA-based interactive patient decision aids. This conclusion was based on the co-development and co-delivery, with two teams of clinicians from the Spanish NHS, of a PDA for the clinical management of Stage IIIA3 (TNM stage T2N2M0) non-small cell lung cancer patients using the templates *Expert Choice* and *Annalisa in Elicia* (*ALEL*).

These two versions of the lung cancer PDAs, although based on large hierarchies of decision criteria, were successfully developed and delivered, as proof-of-concept, in hypothetical consultations replicating actual clinical consultations.

However, the major but not unsuspected finding was that the use by clinicians of alternative templates implementing dissimilar MCDA methods was heavily influenced by the resource constraints inherent to day-to-day clinical practice. There is a fairly direct relationship between higher levels of decision effectiveness (DE) – a term which is defined here as the achievement of normative standards in the resulting PDA produced using a specific template, and higher levels of decision resources (DR), such as the time and cognitive effort required to deliver the PDA produced using that template. The choice by clinicians between 1) templates with higher DE/ higher DR (such as *Logical Decisions*, *HiView*, or *Expert Choice*) and MCDA templates with lower DE/ lower DR (such as *ALEL*) may require clinicians to trade off DE with DR. Decision Resources-Decision Effectiveness (DRDEA) Analysis – a simple transposition of multi-criteria Cost-Effectiveness Analysis to the decision process - proposes that this choice can be characterised as a ‘meta-decision’ of how to decide to decide. It can therefore be made using a meta-multi-criteria decision model (meta-MCDM) comprising a number of relevant DE and DR criteria.

To explore the trade-offs made by clinicians between DE and DR in day-to-day clinical practice, a DRDEA Multi-attribute Value Theory (MAVT) meta-MCDM was developed, as proof of concept, using the MCDA template HiView 3 and was solved by three clinicians in the Spanish NHS in the context of a choice between the *Expert Choice* and *ALEL* templates for developing a PDA for the clinical management of Stage IIIA3 non-small cell lung cancer. The application of the resulting meta-MCDM with each of the three clinicians from hospital H3 in the Spanish NHS showed that the choice between *Expert Choice* and *ALEL* for developing a non-small cell lung cancer PDA did depend on the preferences of the clinicians for trading off 1) decision effectiveness (DE), i.e. the extent to which *Expert Choice* (*ALEL*) facilitates normativity in the model structuring, the preference elicitation, and the evidence generation/ representation incorporated in the eventual PDA, in relation to the decision resources (DR), i.e. the practicality, ease of use, and acceptability to the organisation of an E*xpert Choice* (*ALEL*)-based PDA.

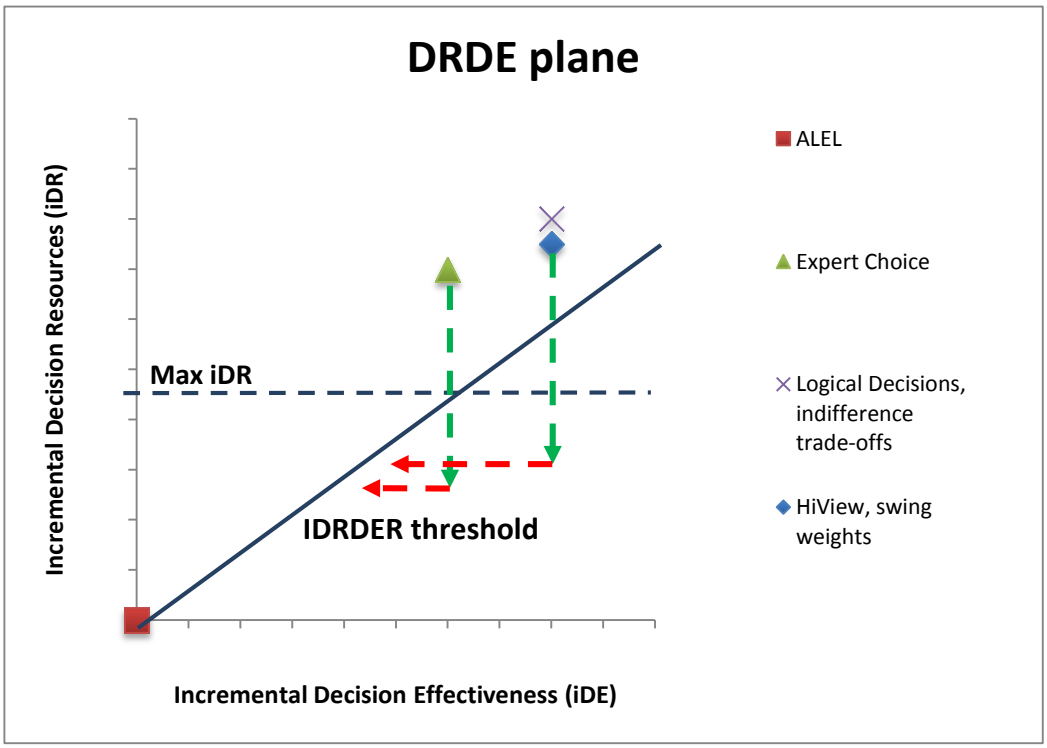
Developing the insights provided by this result we can suggest that clinicians seeking to use MCDA templates with higher DE (that is, ones that *facilitate* the achievement of normative standards in the resulting PDA) will confront a ceiling of DR and that this ceiling may be too high for some templates to be implementable in their full functionality in day-to-day clinical practice. In this research project this was evident in the use of *Expert Choice* to develop a PDA with the team of clinicians from hospital H1. To result in a practical, easy to use, and acceptable PDA, the original hierarchy of decision criteria had to be collapsed into its top-level criteria (effectively paralleling the non-hierarchical template *ALEL*). The inclusion of the entire hierarchy, instead of only its top-level criteria, in the PDA would have accrued a higher level of DE in the use of *Expert Choice*. Perhaps not so much in terms of the normativity of the decision model structure (which was not necessarily well-constructed according to MAVT/MAUT standards – the aim of clinicians being to build a model structure that included as many decision criteria as were thought to be relevant for the decision problem), but certainly in terms of normativity in the preference elicitation and the evidence generation/ representation. With respect to the first, the inclusion of the hierarchy in the PDA would have allowed patients to express their preferences via pair-wise comparisons of relative importance (a preference elicitation technique which generates preferences with reasonable logical

consistency and reasonable empirical accuracy) for all pairs of relevant criteria of the decision from the bottom-level upwards, rather than just for the top-level criteria. With respect to the second, the PDA would have included the evidence of the different options with respect to the wider range of bottom-level decision criteria. Neither of these potential gains in DE was realisable within the DR threshold (i.e. the time constraints of a standard clinical consultation) in the research setting.

Conceptually, the use of a higher DE template to develop a PDA with a DR threshold would only be feasible by sacrificing DE *unless* clinicians were able to operationalize the maximum DE possible within a level of DR that is below the DR threshold. This would require clinicians to build a MCDM structure (i.e. a hierarchy of decision criteria) with the highest possible level of "best practice" or "normativity" in the model structuring achievable by the template which, while containing all the relevant criteria for the patient to make the clinical decision, would be of a small enough size to result in a PDA that can be deliverable within the duration of a clinical consultation, easy enough to use by patients, and acceptable to other clinicians.

In terms of the MAUT formulation of DRDEA, these two approaches to using a high DE template – one involving a sacrifice in DE, one not - to develop a PDA with a DR threshold can be represented in Figures 5.1 and 5.2 – a DRDEA version of the standard Cost-Effectiveness plane.

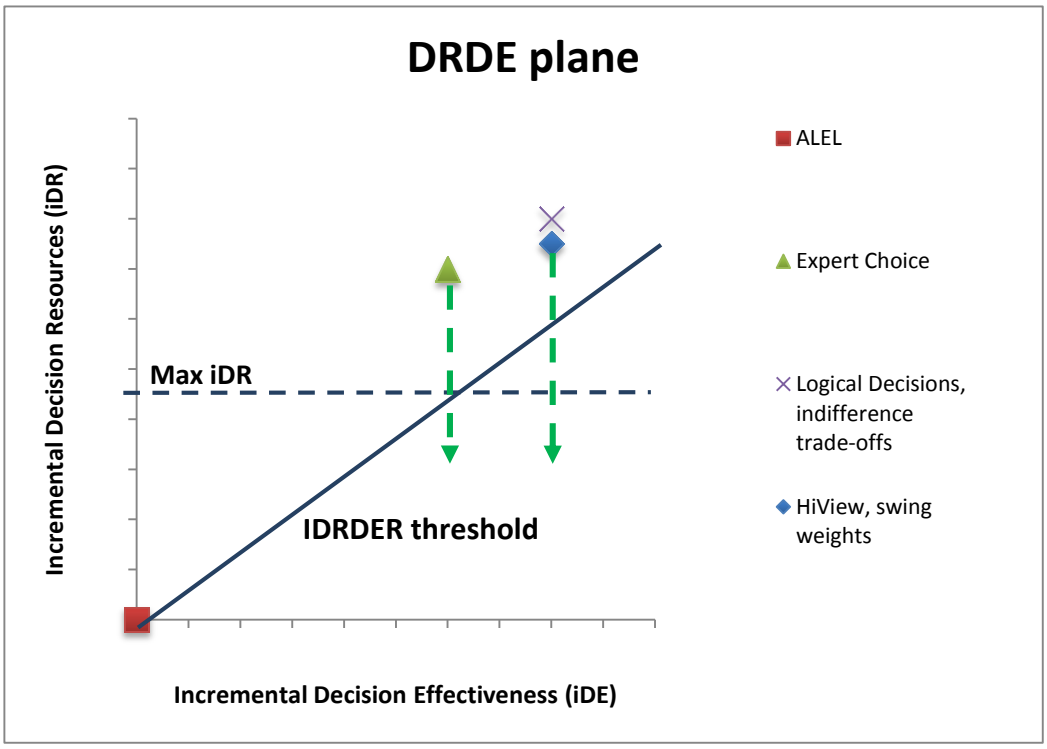
Figure 5.1. DRDEA MAUT formulation. Using high DE templates to develop a PDA: sacrifice in DE



In Figure 5.1, the existence of a ceiling of DR in the use of alternative templates to develop a PDA in the context of clinical practice is represented by a maximum possible level of incremental DR (iDR) in switching from the potentially lowest DE/ lowest DR template (*ALEL*) to a higher DE/ higher DR template (the dashed line marked "Max iDR"). Below the maximum possible level of iDR, the switch from *ALEL* to a higher DE MCDA template will depend on the IDRDER (Incremental DRDE Ratio) threshold (the blue line in Figure 5.1), the slope of which represents the maximum level of additional DR that clinicians consider worth investing per unit of DE gained by making that switch. For MCDA templates below the IDRDER threshold, the switch will be DR-effective. For MCDA templates above the IDRDER threshold, the switch will not be DR-effective. Above the maximum possible level of iDR, the switch from *ALEL* to a higher level DE template is not justifiable. It will only be possible if the iDR required to make that switch is reduced to a level that is lower than the maximum possible level of iDR (as shown in Figure 5.1 with the green arrows). This requires a sacrifice in DE (as shown in Figure 5.1 using the red arrows). For the higher DE MCDA template to be DR-effective, the sacrifice in DE

has to be small enough to be worth the investment in DR made by switching from a lower DE MCDA template (such as *ALEL*) to that the higher DE template.

Figure 5.2. DRDEA MAUT formulation. Using high DE templates to develop a PDA: no sacrifice in DE



In Figure 5.2, the switch from *ALEL* to a higher DE template would be possible without incurring in an iDR which is higher than the maximum possible level of iDR. This is because clinicians would "save" DR by being able to operationalize, for each of the higher DE templates shown in Figure 5.2, the maximum DE possible accruable by the MCDA template.

In the specific context of this research, sacrifices in DE were required. This was primarily due to the large number of criteria (twenty-four) which the clinicians in both hospital H1 and hospital H2 considered were relevant for the decision. The model structure built by the team of clinicians that used the high DE template *Expert Choice* (the team from hospital H1) was too large to result in a PDA with a level of DR below the DR threshold. Although, admittedly, this MCDM structure was not built by the clinicians within an explicit DRDEA framework and could potentially be

reduced by combining some criteria, this was appropriate to the comparative evaluation.

How can templates with the highest possible DE be used to develop and implement MCDA-based PDAs in routine clinical practice with the absolute minimum sacrifice in DR? One answer may lie in the work by Edwards and Barron [121], who developed the Simple Multiattribute Rating Technique with Swings (SMARTS) and SMART Exploiting Ranks (SMARTER) partly in response to the difficulty of the indifference judgments required from the decision maker in order to construct multiattribute preference structures using the best practice standards of MAVT/MAUT proposed by Keeney and Raiffa [11]. Edwards and Barron suggested the use of a "strategy of heroic approximation" [121] to identify "the simplest possible judgments that have any hope of meeting the underlying requirements of multiattribute utility measurement, and try to determine whether they will lead to substantial suboptimal choices in the problem at hand" [121]. If they do not lead to suboptimal choices (which Edwards and Barron suggest in many decision situations will be the case), a quick and easy to implement nine-step procedure (see Chapter 1, section 8.2.1.1) can be used to solve a multi-criterion decision problems according to prescriptive standards. Templates such as Logical Decisions and HiView 3 may be used to implement SMARTS as the basis of MAVT-based PDAs in routine clinical practice. Templates such as Logical Decisions and Web-HIPRE may be used to implement SMARTER for the same task.

2. Study limitations.

The potential weaknesses of the methods used in this research project need to be highlighted. The project was undertaken in close collaboration with clinicians in the Spanish NHS with a view to understanding how the use of MCDA in developing and delivering interactive patient decision aids is affected by the actual context of day-to-day clinical practice. For that reason, the methods were context-led, that is, they were adapted to the clinicians' point of views and daily routines during the course of the study which partook of many of the qualities of 'action research'. In this sense the clinicians' preferences with regards to the content of the *Expert Choice*-based PDA and the ALEL-based PDAs, as well as in regard to the inputs of the meta-decision model, were accepted independently of any considerations of methodological rigour,

as were the clinicians' strong time constraints on participation in this research. The potential weaknesses of the methods need to be interpreted within this necessarily context-led approach.

2.1. Limitations in the development and delivery of the Expert Choice-based and the ALEL-based M-IPDAs.

2.1.1. The hierarchy of criteria for the Expert Choice-based and the ALEL-based PDAs

In both hospital H1 and hospital H2, the processes of building the final hierarchy of decision criteria was researcher- and clinician-led and did not seek to adopt the patient-centred perspective with patient involvement in the determination of patient-important outcomes and criteria.

The final set of decision criteria was comprehensive as far as the clinicians were concerned. In terms of model structuring, however, this led to a decision hierarchy which was not, strictly speaking, well-structured according to normative standards of decision theory. In particular, several bottom-level criteria were arguably not preference independent. In this case, strict normativity was sacrificed to the clinicians' desire for subjective comprehensiveness in the model structuring. This can be reasonably justified from a practical point of view. The lack of value independence does not necessarily invalidate the use of value dependent decision criteria in MCDA [251]. However, it could have been desirable to engage with the clinicians in a requisite decision modelling [212] exercise in this respect.

Such an exercise would have consisted in a consultative and iterative process between the clinicians and the present author with the aim of modifying the model structure to be 1) more in line with the normative axioms of decision theory while containing 2) the relevant content to the clinicians for helping patients make the decision. However such a modelling exercise was not possible in either hospital H1 or hospital H2 due to time constraints on the side of the clinicians derived from their clinical duties.

2.1.2. Generating the evidence

For both the Expert Choice-based and the ALEL-based PDA, the evidence of the performance of the lung cancer clinical management options on the bottom-level decision criteria was generated in a highly pragmatic way that respected the time constraints of the clinicians involved in this research - eliciting their expert opinion by asking them to state their estimates of the options' performance levels. Furthermore, for all the quality of life criteria, the estimates of the performance levels of the decision alternatives were limited to one point in time – 2 years after the start of treatment.

The performance levels resulting from the elicitation of the clinicians' expert opinion differed, in some cases substantially, across clinicians. There are at least three possible reasons for these differences:

1) The large uncertainty inherent to the raw performance levels being assessed from the clinicians, as these raw performance levels depend on a series of probabilistic events. For example, the level of pain that a NSCLC patient is expected to have two years after starting treatment with neoadjuvant chemotherapy depends on his health state 2 years after starting treatment which depends, among other probabilities, on the probability that 1) the neoadjuvant chemotherapy treatment has been successful at downstaging the tumour, 2) the probability that the surgical resection is complete, and 3) the probability that, if the surgical resection is not complete, after further chemo-radiotherapy there is a local or advanced recurrence;

2) Differing clinician beliefs. There is evidence in the lung cancer literature that lung cancer clinicians express beliefs (i.e. subjective probabilities) regarding the outcomes of lung cancer treatments that differ substantially from both their peers and with respect to the published evidence [252], an argument which resonates with the results of other research suggesting that in general experts predict poorly [253];

3) Lack of accuracy of the methods used in this research project to obtain the raw performance levels.

With more time and financial resources (unavailable to the clinicians in this research project), clinicians could attempt to improve the quality of the estimates of the

evidence by undertaking a review of the scientific literature regarding the impact of each treatment option on each bottom-level decision criterion of the Expert Choice-based and the ALEL-based PDAs. The aim of these literature reviews would be to find the best available published evidence of the performance levels, from the start of treatment until death, of the clinical management options on each of these criteria. The best available published evidence should come from high quality sources (e.g. clinical trials, systematic reviews and/or meta-analyses), but it is vital to emphasise that this evidence is likely to be unavailable for many of the decision criteria and for many individualised cases within a condition. A review of the medical literature undertaken by the present author revealed no studies comparing, head to head, the lifetime clinical (or other) outcomes of the three non-small cell lung cancer treatment options (neoadjuvant chemotherapy, concurrent chemo-radiotherapy and best supportive care). Some studies evaluating the outcomes of these options were found, but they generally related to patients in more than one stage of non-small cell lung cancer, and often to interventions that are dissimilar to the three clinical management options considered in this research project.

Best practice modelling by 'practice –normative' standards requires clinicians to input into the decision model the "best estimates available now" (BEANs) for the raw performance levels on the relevant decision criteria. The BEANs will most likely come from expert opinion elicitation, although a few may be available from published scientific papers based on higher quality sources. It is of course arguable that the expert judgments of clinicians are made in the light of familiarity with this literature as modified by their clinical experience. Elicitation of expert opinion should be undertaken using best practice methods such as those described in von Winterfeldt and Edwards [151].

2.1.3. Format and content of the Expert Choice-based and the ALEL-based PDAs

Neither of the two PDAs was developed to achieve all of the best practice standards of format and content required by the International Patient Decision Aids Standards [67]. Notwithstanding the criticisms made to these standards by, among others, McDonald and Charles [89] and Bekker [87], the main reason for not focusing on these standards to develop the Expert Choice-based and the ALEL-based PDAs was

that, in this particular research project, the objective was not to develop two PDAs that met abstract external standards of format and content quality. Rather, it was to develop two PDAs that could potentially improve the quality of patient decision making in clinical consultations with respect to current practice (that is, with respect to verbal reasoning). This is not to say that it is not desirable to develop these two tools with reference to external standards as targets, but not as evaluation checklists as such.

2.1.4. Possible bias in the delivery of the M-IPDAs in clinical consultations

With respect to the delivery of the two PDAs, due to time constraints, neither the clinicians nor the proxy patients had training in the use of the two templates. In fact, the hypothetical consultations were followed using a script and strong support from the researcher. This is likely to have introduced a bias in the interaction between the clinician and the proxy patient. It certainly did not allow for a proper assessment of the fluidity of the communication between clinicians and proxy patients using the PDAs.

2.1.5. Proxy patients versus real patients in the delivery of the M-IPDAs

Perhaps the greatest limitation in the delivery of the two MCDA-based PDAs was the use of proxy patients and hypothetical consultations instead of real patients and real consultations. There is a large gap between hypothetical consultations and real consultations. Proxy cancer patients are very different from real cancer patients. Anxiety, a frequent response to cancer diagnosis [254] is likely to interfere with the patient's ability to communicate with his/her physician to discuss the information necessary to make a balanced treatment decision. Although there is some evidence that the exposure to PDAs does not increase anxiety in patients [255, 256] it is likely that actual lung cancer patients, especially if they have relatively low educational and computer-literacy levels, will find, within a generally anxious state of mind, the use of the PDAs more challenging than proxy patients. Nevertheless it is important not to stereotype or pre-judge individual patients in this respect and an important step to understand the impact of delivering the Expert Choice-based PDA and the ALEL-based PDA in the context of clinical practice is to pilot the use of these tools in actual day-to-day consultations with real Stage $IIIA_3$ non-small cell lung cancer patients.

Given the heterogeneity of patients, the average results from such piloting need careful interpretation.

2.2. Limitations in the development and delivery of the DRDEA meta- decision model.

2.2.1. The scope of the DRDEA meta-decision model.

The main limitation of the meta-MCDM was its limited scope. In total, 11 decision effectiveness (DE) and decision resources (DR) criteria were included in the model. Although the inclusion of 11 criteria was justified by the 1) proof-of-concept approach and 2) the time constraints of clinicians in hospital H3 to participate in this research project, the meta-decision model can be extended to include other important DE and DR criteria. In terms of DE, for example, an important criterion for inclusion in further versions of the meta-decision model is the DE in the delivery of the PDA, e.g. the extent to which the use of a template in the delivery of a PDA increases the quality of the decision. In term of DR, an important criterion for inclusion in further versions of the meta-decision model is the practicality in the development of a PDA, which can be defined as the extent to which it is practical to develop a PDA based on a particular template within the time and organisational constraints available to clinicians for that task. The development of a PDA can require a large amount of time and other resources if, for example, in order to build a well-structured decision hierarchy, clinicians hire a decision analyst. Or if clinicians engage in time-consuming reviews of the literature and/or modelling exercises to generate the evidence of the clinical options on the decision criteria.

Two other types of DE and DR criteria that could also be considered for inclusion in future versions of the meta-decision model are 1) in relation to DE, the improved health outcomes associated with using a particular template to deliver a PDA and 2) in relation to DR, the changes in health care expenditures, associated with using a template to deliver a PDA.

With respect to the first of these two criteria, several studies have evaluated the actual impact of (non-MCDA-based) patient decision aids on health outcomes. For example, Barry et al [257] in a randomised controlled trial evaluating the impact of a multimedia decision aid on the risks and benefits of different treatments for benign

prostatic hyperplasia, measured urinary symptoms at 3 months after the use of the decision aid via a urinary symptoms index measure. In another study, Deyo et al. [258] carried out a trial of the impact on an array of outcomes of an interactive video disk giving information about alternative treatments for back injuries. The authors measured, among others, back pain severity at one year after patients were exposed to the decision aid. They found a statistically significant effect (a reduction in back pain severity) of the intervention compared with the control (no videodisk). Murray et al [259] conducted a clinical trial to evaluate the impact of 1) a clinical decision aid consisting of an interactive multimedia program accompanied by a booklet describing hormone replacement therapy risks and benefits versus 2) no clinical decision aid on, among other outcomes, self-assessed health status using both the EuroQoL EQ-5D quality of life questionnaire and the SF-36 quality of life questionnaire. 205 menopausal women were randomised to the intervention and control. Health-related quality of life was assessed at 3 and 9 months after using the decision aid. The authors found no significant changes in the baseline health status scores over time in either of the two groups. Finally, Kennedy et al [104] used the SF-36 questionnaire to evaluate the impact of two decision aid tools, 1) a booklet and accompanying videotape and 2) the same booklet and videotape plus a preference elicitation interview- versus 3) no decision aid on the quality of life of women with uncomplicated menorrhagia. 894 women were randomly allocated across the three groups, and asked to fill in the SF-36 questionnaire at 2-year follow-up. At this time, there was a statistically significant difference between group 2) and the other 2 groups in the SF-36 role function score.

With respect to the second criterion, the impact of introducing patient decision aids on health service costs has been evaluated albeit in too few studies to draw substantive conclusions. To mention a study which achieved positive results, Kennedy et al [104] performed a cost analysis of the impact of using 1) a booklet + videotape and 2) booklet + videotape + preference elicitation interview as decision aids for women with menorrhagia. The cost perspective was the UK NHS perspective. The authors calculated the resources used during the development and production of the interventions, including 1) the duration of time devoted by nurses during the preference elicitation technique (if any), 2) use by the women in the study of health services, including tests and other procedures, medications for menorrhagia,

inpatient days in hospital and outpatient and family physician visits (if any) at 6, 12, and 24 months after the interventions. The effective life of the interventions was estimated at 3 years. Costs per patient were calculated by dividing the total cost of each intervention by the number of patients in each intervention group. Both intervention groups showed major mean cost savings compared to the control group.

2.2.2. Evaluating the performance of alternative MCDA templates on the criteria of the meta-decision model.

One of the main concerns that arose in the development of the meta-decision model was providing the estimates of the performance levels of the proposed PDAs on the meta-decision criteria. Because making the decision of what type of template to use in the development and delivery of a PDA has to be made *ex ante*, it is not possible to accurately measure many of these performance levels – e.g. the impact of using *this* MCDA template on criteria of DE in the delivery of *this* M-IPDA. This is of course a universal problem in decision making and decision support and one that is simply made more obvious in, but not created by, developing an meta-decision model. One productive route would involve moving from evaluating the templates as such to evaluating their use in particular ways and within particular constraints. For example one option in this more refined meta-model might be: "*Expert Choice* being used to produce a PDA that can be delivered in less than 1 hour, is easy to use by patients after approximately 30 minutes of training and that 90% of clinicians in the hospital will consider suitable for use in their clinical consultations".

Using the "best evidence available now" (BEANs) standard, in order to learn the extent of DE accrued and DR consumed developing and delivering PDAs, information is required from the BEANs. Currently, there is little information in the literature about the impact of PDAs produced using the different templates reviewed in this research in terms of practicality in the preference elicitation or organisational fit, but hopefully this will change in the future as the use of templates to develop and deliver MCDA-based PDAs becomes more widespread. The BEANs will then be based on more robust study designs. In the absence of such evidence from the literature, the performance of alternative templates in terms of DE and DR needs to be estimated using best practice expert opinion elicitation methods as in the decision

modelling itself. The need to make a meta-decision can no more be postponed than the decision itself.

3. Generalizability of results and implications for public health.

The work undertaken in this study is a proof-of-concept study exploring 1) the potential of using MCDA in the development and implementation of PDAs in routine clinical practice and 2) the potential for the application of the Decision Resources-Decision Effectiveness Analysis (DRDEA) MCDA framework to the choice by clinicians of MCDA template as the basis for a PDA. The context in which this proof-of-concept study was carried out was that of Spanish NHS tertiary hospitals, so the generalisability of the results to other health care settings is limited to hospitals with similar characteristics. The teams of clinicians who participated in this study were selected by this researcher based on prior experience doing research together. They are all highly specialised, highly motivated clinicians. In this sense, they may not be representative of the population of all clinicians.

The implementation of the *Expert Choice*-based PDA and the *ALEL*-based PDA in hypothetical consultations replicating actual clinical consultations was done with proxy patients who were healthy individuals. The results of these consultations cannot be generalised to actual lung cancer patients.

In terms of generalisability to other therapeutic areas, the decision of what is the best clinical management strategy for a Stage $IIIA_3$ non-small cell lung cancer patient has several particular characteristics. First of all, it is a very complex decision problem. The options are composed of uncertain interventions which vary depending on the evolution of the patient. Both the condition and the clinical management options affect many aspects of the patient's well-being. The decision is fateful, so the PDA delivery is likely to cause high level of anxiety to the patients. In many other therapeutic areas, the decision problem is likely to be more straightforward and the applicability of MCDA easier. It is in this sense that the methods used to develop and deliver the Expert Choice-based PDA and the ALEL-based PDA with clinicians in the Spanish NHS is likely to be applicable to many other clinical decisions. With respect to the trade-offs between decision resources and decision effectiveness, it is also likely that for other therapeutic areas in which the decision problem is easier to

conceptualise (e.g. clearer options, less relevant criteria) these trade-offs will be less important, facilitating the use of templates which implement high DE MCDA methods (such as MAVT).

In terms of the public health implications of this research, Stage $IIIA_3$ non-small cell lung cancer is a rare condition. Implementing MCDA-based PDAs for the clinical management of this condition is not likely to have much importance to public health in terms of burden of disease. However, the active treatment options for this condition are very expensive. Allowing patients to make values-based choices in this realm is likely to have cost implications. For example, if more patients choose, based on their values, best supportive care than they would having not been exposed to the PDA. One of the original motivations for the development of PDAs was to help reduce unwarranted variations in the provision of health care (i.e. variations in the provision of care that could not be explained by the need for these interventions). In this sense, PDAs aim to spur patient self-interest in avoiding such interventions [23]. The clinical management of Stage $IIIA_3$ is a case in point: patients with strong preferences for avoiding the many adverse effects of the active treatments may decide that it is best for them to receive best supportive care.

4. Suggestions for further research.

1) There is a need for more research on the use of high DE, low DR templates to develop PDAs with clinicians using prescriptive MCDA approaches such as SMARTS/SMARTER;

2) There is a need for more research to identify best practice approaches to delivering prescriptive MCDA-based PDAs to patients;

3) There is a need for more research to identify best practice methods for developing and delivering prescriptive MCDA-based PDAs for clinical management decisions in oncology within the constraints imposed by the decision resources relevant for the particular decision at hand. Decisions about the clinical management of oncological conditions can be fateful, are often characterised by uncertainty, and often result in recurrence of disease. How should patients be supported to make these decisions using prescriptive MCDA considering the levels of anxiety that patients can be in? How can uncertainty be modelled into these PDAs so that it is understood by

patients? When should these PDAs be delivered? In oncology, there are particular points of time in the evolution of the disease whe MCDA might be more appropriate than at other moments;

4) What is the role of decision analysts in supporting both the development of prescriptive MCDA-based PDAs with clinicians and their implementation with patients? This role is likely to vary with the particular condition under consideration. For less fateful, less uncertain, less complex decisions with few decision criteria decision analysts may have a less crucial role than for fateful, uncertain, complex decisions;

5) What are the most appropriate decision support software interfaces to provide decision support for prescriptive MCDA-based PDAs? These software applications should implement prescriptive MCDA approaches with the highest possible level of decision effectiveness but with the lowest possible requirements in terms of time and cognitive effort.

Appendix 1: Initial hierarchies of criteria for the *Expert Choice*-based PDA and the *ALEL*-based PDA.

Figure A1.1. Initial hierarchy of criteria for the Expert Choice-based PDA

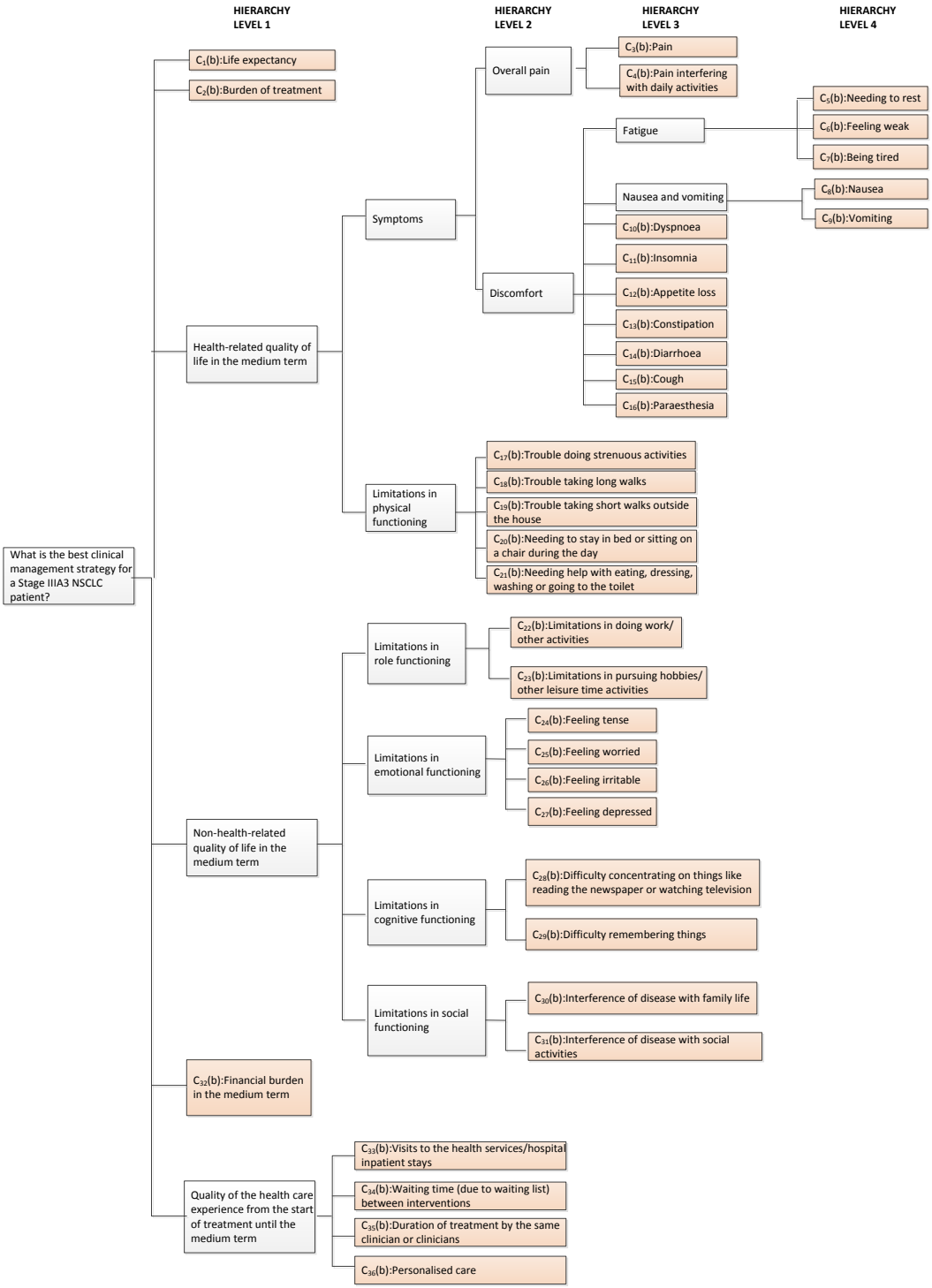
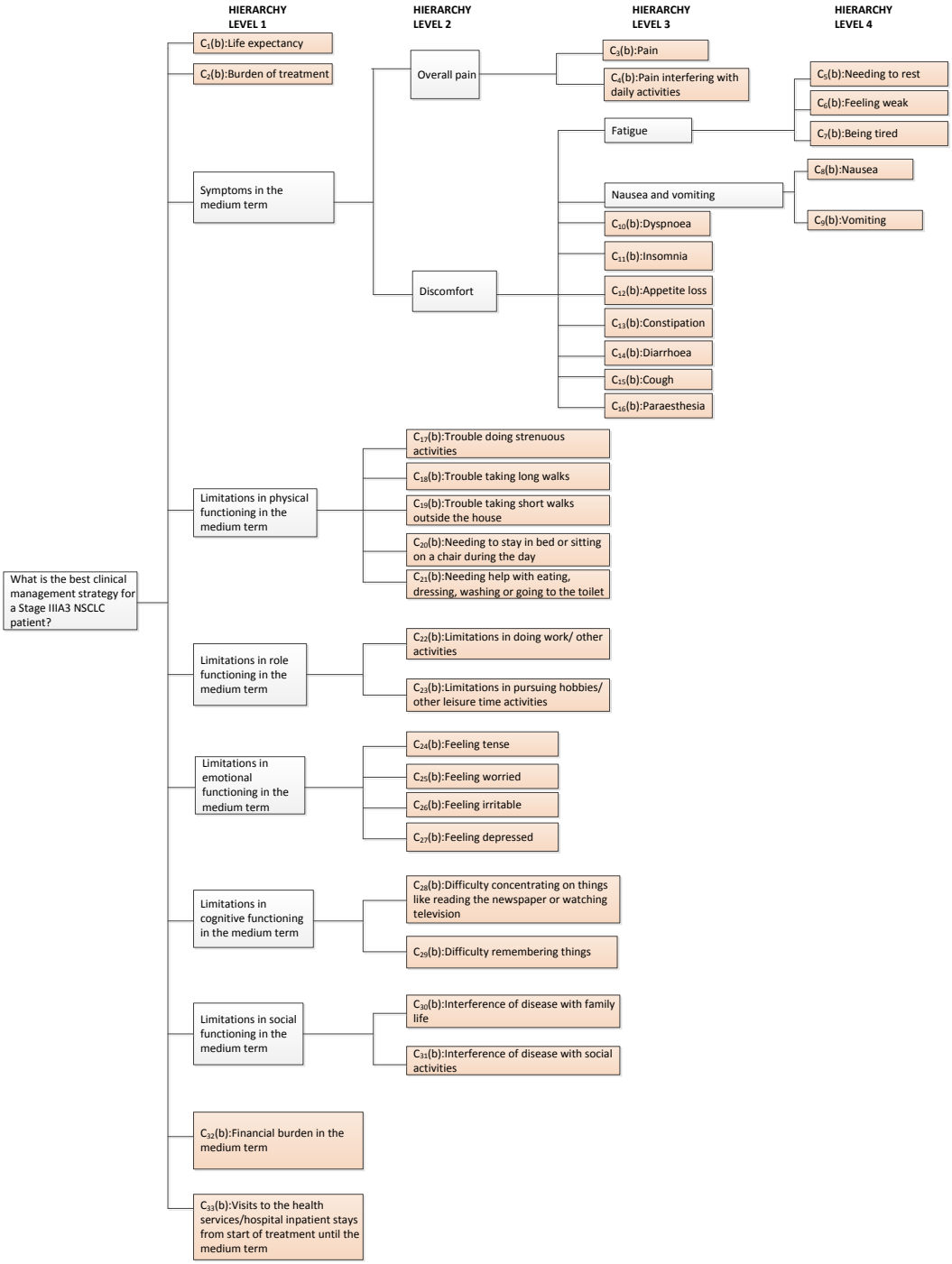


Figure A1.2. Initial hierarchy of criteria for the ALEL-based PDA



Appendix 2: Description of the criteria of the final hierarchy common to the *Expert Choice*-based PDA and the *ALEL*-based PDA.

The final hierarchy of criteria was common to the *Expert Choice*-based PDA and the *ALEL*-based PDA. The six top-level criteria of this common hierarchy were:

1. Cure from cancer, i.e. how likely it is to get cured from Stage $IIIA_3$ NSCLC

2. The duration of life, i.e. the life expectancy

3. The quality of life in the medium term, where medium term is defined as two years after the start of treatment

4. The disease-related financial burden in the medium term, i.e. the financial problems derived, two years after the start of treatment, from 1) direct expenditures related with the disease and/or the treatment, and from 2) the opportunity cost of not being able to earn a living as a result of being ill

5. The treatment-related adverse effects

6. The quality of the health care experience (i.e. those aspects of the health care delivery which are positive for one's well-being as a patient) from the start of treatment until the medium term.

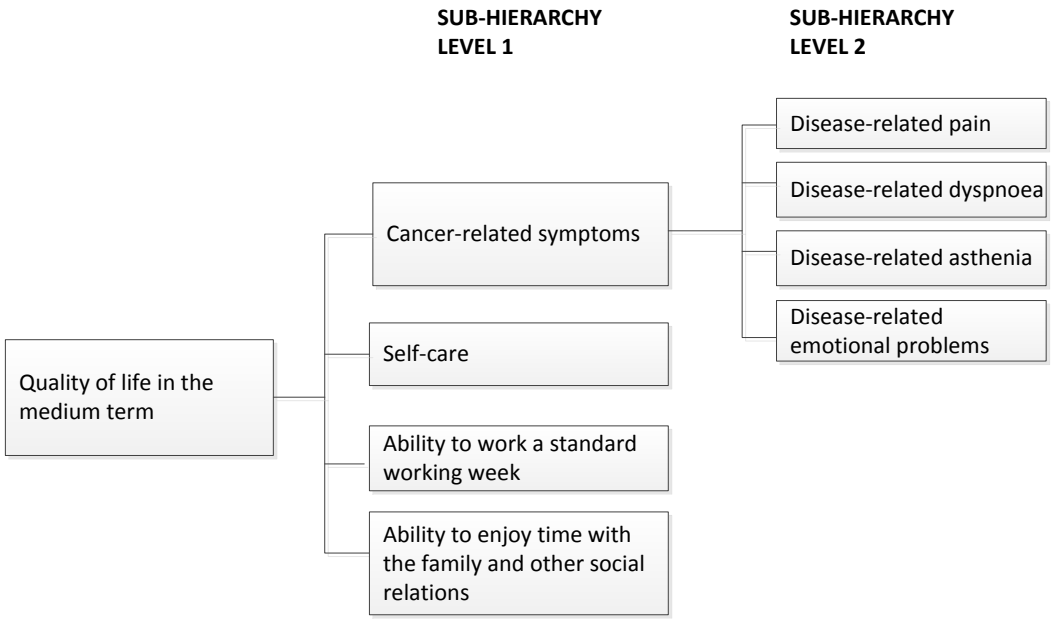
Note that, compared to the initial set of decision-relevant aspects from Appendix 1, an additional aspect was included in the final decision hierarchy: cure from cancer. This criterion was added after informal conversations between the present author and two of the participating clinicians. In these conversations, held after the two model-building exercises, both clinicians pointed out that it was indeed possible (albeit unlikely) for the hypothetical Stage $IIIA_3$ non-small cell lung cancer patient to get cured if the patient chose to undergo one of the two active treatment strategies, particularly option A_1 (neo-adjuvant chemotherapy with respective intent). Cure from Stage $IIIA_3$ NSCLC was not an easy concept to define for the hypothetical patient, as there is always the chance that metastatic lesions may recur. In this project, the clinicians agreed to define cure from cancer as the absence of any tumour activity after 5 years of starting the treatment.

Three top-level criteria (cure for cancer, duration of life, and disease-related financial burden in the medium term) were defined as stand-alone criteria. The three remaining top-level criteria (quality of life in the medium term, treatment-related adverse effects, and quality of the health care experience from the start of treatment until the medium term) were defined in terms of sub-criteria. These are explained below.

A2.1. Defining quality of life in the medium term in terms of a sub-hierarchy of sub-criteria

In the final hierarchy of criteria common to both the *Expert Choice*-based PDA and the *ALEL*-based PDA, quality of life in the medium term (i.e. after 2 years of starting treatment) was defined in terms of a sub-hierarchy of sub-criteria. Level 1 of the sub-hierarchy was defined by four sub-criteria characterising quality of life in the medium term: 1) cancer-related symptoms, 2) self-care (i.e. being able to take care of oneself without help from others), 3) being able to work a standard working week (i.e. 40 hours), 4) interference of the disease with family life and/or other social activities. Level 2 of the sub- hierarchy was composed of four cancer-related symptoms children criteria. These sub-criteria were considered to be by the two teams of clinicians the symptoms due to the disease that most commonly occur in Stage $IIIA_3$ NSCLC patients in the medium term. They were: 1) disease-related pain, 2) disease-related dyspnoea (difficulty breathing), 3) disease-related asthenia (feeling of weakness), and 4) disease-related emotional problems (i.e. anxiety and/or depression). Figure A2.1 shows this sub-hierarchy of quality of life in the medium term sub-criteria.

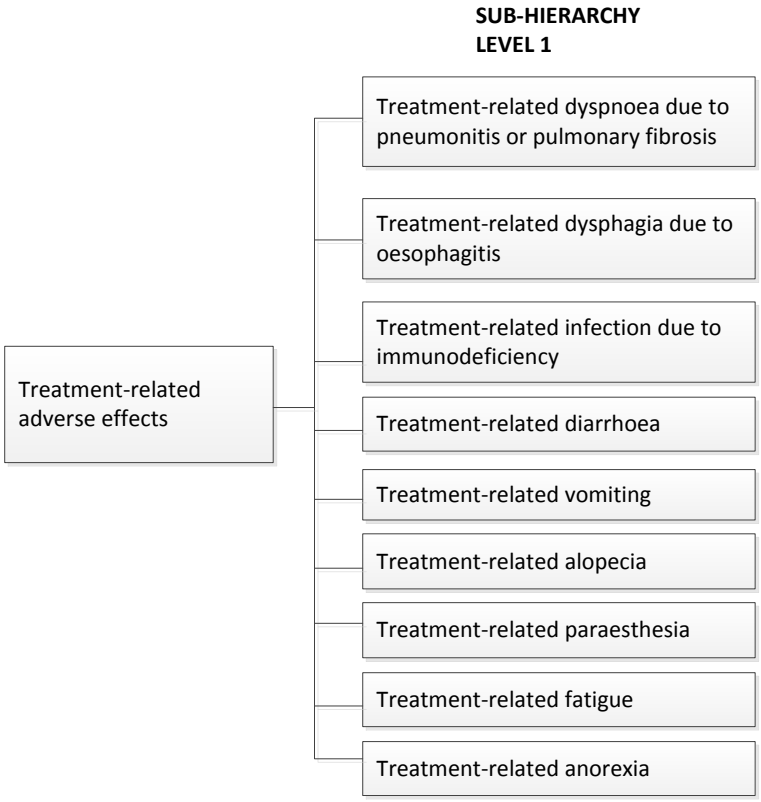
Figure A2.1. Sub-hierarchy of quality of life in the medium term sub-criteria (Expert Choice-based PDA and ALEL-based PDA)



A2.2. Defining treatment-related adverse effects in terms of a sub-hierarchy of sub-criteria

The treatment-related adverse effects criterion was defined as a one-level sub-hierarchy of sub-criteria. Deciding what these sub-criteria should be was not an easy task, as more than five hundred cancer treatment-related adverse effects have been described in the literature (these are detailed and graded, in terms of severity, in the National Cancer Institute Common Terminology Criteria for Adverse Effects (CTCAE) grading system). For practical reasons, the approach taken by this researcher to select these sub-criteria was to reach a consensus with both teams of clinicians about which were the nine most common adverse effects associated with the treatment of the Stage $IIIA_3$ NSCLC patient. The nine treatment-related adverse effects sub-criteria were 1) dyspnoea as a consequence of pneumonitis or pulmonary fibrosis, 2) dysphagia (i.e. problems swallowing) as a consequence of oesophagitis, 3) infection as a consequence of immunodeficiency, 4) diarrhoea, 5) vomiting, 6) alopecia (hair loss), 7) paraesthesia (feelings of tickling, burning, or numbness in the skin), 8) fatigue, and 9) anorexia (loss of appetite). Figure A2.2 shows the sub-hierarchy.

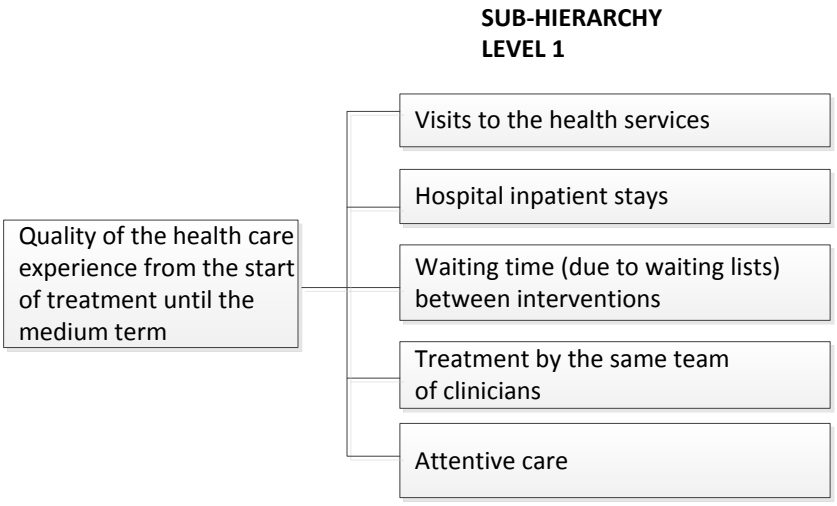
Figure A2.2. Sub-hierarchy of treatment-related adverse effects sub-criteria (Expert Choice-based PDA and ALEL-based PDA)



A2.3. Defining quality of the health care experience from the start of treatment until the medium term in terms of a sub-hierarchy of sub-criteria

This criterion was defined as a one-level sub-hierarchy of sub-criteria. The chosen sub-criteria were 1) visits to the health services, 2) hospital in-patient stays, 3) waiting time (due to waiting lists) between interventions, 4) treatment by the same team of clinicians (that is, whether or not the same clinician or clinicians follow-up the patient throughout time), 5) attentive care (that is, whether or not the patient is always treated by his clinicians in a considerate and caring fashion). Figure A2.3 shows the sub-hierarchy.

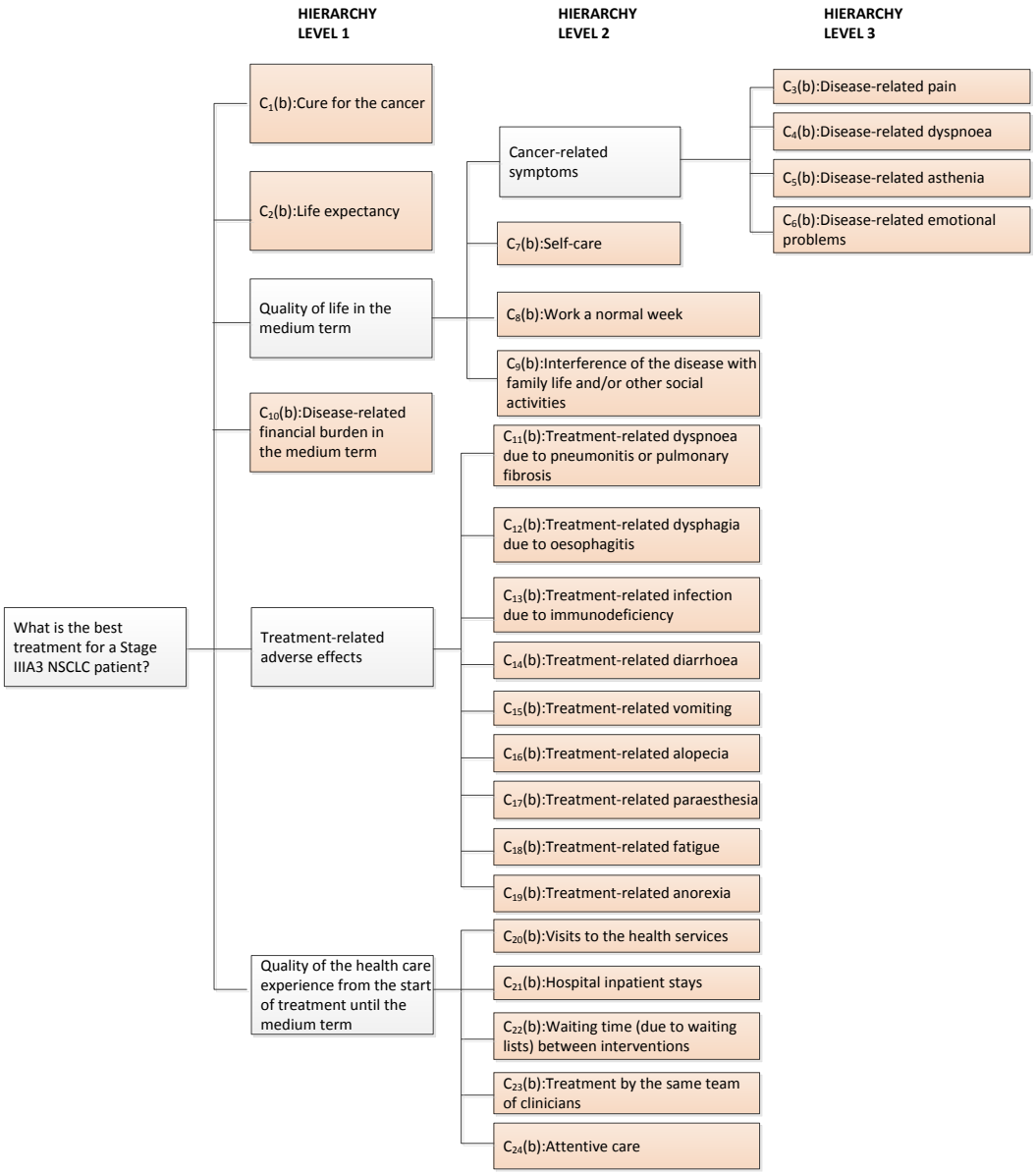
Figure A2.3. Sub-hierarchy of health care experience between the start of treatment and the medium term sub-criteria (Expert Choice-based PDA and ALEL-based PDA)



Appendix 3: Literature search of published clinical studies evaluating the outcomes of the three clinical decision options.

The purpose of this literature search was solely to identify the types of outcomes that are measured in existing clinical studies evaluating the impact of the clinical management strategies used in this research as decision options for both the *Expert Choice* and the *ALEL* Stage $IIIA_3$ NSCLC treatment PDAs, i.e. chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2), and best supportive care (A_3). Identifying these outcomes was important to understand how likely it was that the currently existing evidence regarding the impact of these three decision options on the bottom-level criteria $C_k(b)$ $(k = 1 \ldots 24)$ of the full final decision hierarchy for the two PDAs (shown in Figure A3.1, where the relevant criteria are highlighted in red colour) could be used to inform the calculation of the scores of the three decision options on these twenty-four criteria.

Figure A3.1. Full final hierarchy of criteria for the Expert Choice-based PDA and the ALEL-based PDA



The literature search, which was done in PUBMED. It was designed to identify all the clinical trials, comparative studies, evaluation studies, meta-analyses, reviews and systematic reviews published in English in the last seven years (i.e. since January 1, 2006) which would be specific to patients with stage $IIIA_3$ non-small cell lung cancer over sixty-five years of age and which would include, either as the main treatment or as a comparator, the following options:

1. Neoadjuvant chemotherapy followed by surgery with one of the chemotherapy agents being either cisplatin or carboplatin

2. Concurrent chemotherapy with one of the chemotherapy agents being either cisplatin or carboplatin

3. Best supportive care

The search strategy, including the search terms and the number of published articles associated with each search term are shown in Table A3.1.

Table A3.1: Search strategy

Individual search identification number	**Search terms**	**Number of articles found in the database**
#1	Cancer* OR carcinoma* OR neoplasm*	2,433,263
#2	Non-small cell OR non small cell	232,018
#3	Lung OR pulmonary	952,371
#4	#1 AND #2 AND #3	44,631
#5	T2N2M0 OR IIIA3 OR 3A3 OR IIIA OR 3A OR locally advanced	52,556
#6	#4 AND #5	3,354
#7	Induction chemotherapy OR neoadjuvant chemotherapy	38,656

Table A3.1 (cont.): Search strategy

Individual search identification number	**Search terms**	**Number of articles found in the database**
#8	#6 AND #7	863
#9	Cisplatin OR carboplatin	58,527
#10	#8 AND #9	494
#11	Lobectomy OR surgical resection OR tumour resection OR surgery	44,631
#12	#10 AND #11	303
#13	Limit #12 to: 7 last years; patients aged 65 or older; clinical trials, comparative studies, evaluation studies, meta-analyses, reviews and systematic reviews	40
#14	Concurrent chemo-radiotherapy OR concurrent radio-chemotherapy OR concurrent chemo radiotherapy or concurrent radio chemotherapy OR (chemotherapy AND radiotherapy)	85,259
#15	#9 AND #14	10.523

Table A3.1 (cont.): Search strategy

Individual search identification number	**Search terms**	**Number of articles found in the database**
#17	Limit #16 to: 7 last years; patients aged 65 or older; clinical trials, comparative studies, evaluation studies, meta-analyses, reviews and systematic reviews	123
#18	No active treatment OR best supportive care OR palliative care NOT (chemotherapy OR radiotherapy)	163,383
#19	#6 AND #18	16
#20	Limit #19 to: 7 last years; patients aged 65 or older; clinical trials, comparative studies, evaluation studies, meta-analyses, reviews and systematic reviews	2

Using the search strategy shown in Table A3.1, forty articles were found that potentially described the outcomes of neoadjuvant chemotherapy with resective intent (see search #13) in Stage $IIIA_3$ non-small cell lung cancer patients. In addition, a hundred and twenty-three articles were found which potentially described the outcomes of concurrent chemo-radiotherapy (see search #17) in the same type of patients. Finally, two articles were found which potentially described the outcomes of best supportive care (see search #20) in those patients. The abstracts of all these articles were reviewed. An article was selected for review if, of all the patients treated with the relevant option in the study, the group of patients with Stage $IIIA_3$ non-small cell lung cancer was the most numerous. If it was not possible to know from the article the number of Stage $IIIA_3$ patients treated with the relevant option, the article was selected for review if the number of Stage IIIA patients treated with the relevant option in the study was more than 50% of all the patients described in

the article. The rationale for this was to ensure that the evidence retrieved from the literature referred as much as possible to the same type of patient who would be engaged in decision making with the *Expert Choice* or *ALEL* PDAs – i.e. patients with Stage $IIIA_3$ NSCLC.

With the above criteria for article selection, seven (out of the original forty) articles were found describing the outcomes of neoadjuvant chemotherapy with resective intent in Stage $IIIA_3$ non-small cell lung cancer patients (see Table A3.2). With the same criteria, five articles (out of one hundred and twenty-three) were found describing the outcomes of concurrent chemo-radiotherapy in Stage $IIIA_3$ non-small cell lung cancer patients (see Table A3.3). No articles were found describing the outcomes of best supportive care in Stage $IIIA_3$ non-small cell lung cancer patients.

Table A3.2 presents, for each of the seven articles selected for review regarding the outcomes of neoadjuvant chemotherapy with resective intent (A_1) for Stage $IIIA_3$ NSCLC patients, a brief description of the type of study, the main treatment and comparators described, and the outcomes reported.

Table A3.2: Results of the literature search. Treatment option: neoadjuvant chemotherapy with resective intent (A_1)

Article	**Type of study**	**Main treatment/ comparator (if any)**	**Outcome(s) reported**
Brunelli, A., et al., *Gemcitabine-cisplatin chemotherapy before lung resection: a case-matched analysis of early outcome.* The Annals of Thoracic Surgery, 2006. **81**: p. 1963-1968	Observational study	Main treatment: Neoadjuvant chemotherapy (cisplatin + gemcitabine) with surgery Comparator: Only surgery	Cardiopulmonary morbidity/ mortality/ Perioperative blood transfusion/ emergency ICU admissions/ length of postoperative stay/ bronchopleural fistula/ prolongued air leak/ empyema

Table A3.2 (cont.): Results of the literature search. Treatment option: neoadjuvant chemotherapy with resective intent (A_1)

Article	**Type of study**	**Main treatment/ comparator (if any)**	**Outcome(s) reported**
Esteban, E., J. de Sande, and N. Villanueva, *Cisplatin plus gemcitabine with or without vinorelbine as induction chemotherapy prior to radical locoregional treatment for patients with stage III non-small-cell lung cancer (NSCLC): results of a prospective randomized study*. Lung Cancer, 2007. **55**: p. 173-180.	Clinical trial	Main treatment: Neoadjuvant chemotherapy (cisplatin + gemcitabine) with surgery (if downstaging) or radiotherapy (if no downstaging) Comparator: Neoadjuvant chemotherapy (cisplatin + gemcitabine + vinorelbine) with surgery (if downstaging) or radiotherapy (if no downstaging)	Tumour response/ survival/ stable disease/ disease progression/ anaemia/ neutropenia/ thrombocytopenia/ infection/ fever/ nausea and vomiting/ asthenia/ alopecia
Gottfried, M., R. Ramlau, and M. Krzakowski, *Cisplatin-based three drugs combination (NIP) as induction and adjuvant treatment in locally advanced non-small cell lung cancer*. Journal of Thoracic Oncology, 2008. **3**: p. 152-157.	Clinical trial	Main treatment: Neoadjuvant chemotherapy (cisplatin + vinorelbine + ifosfamide/mesna) with surgery and post-surgical chemotherapy Comparator: Neoadjuvant chemotherapy (cisplatin + vinorelbine + ifosfamide/mesna) with surgery	Tumour response/ survival/ disease progression/ anemia/ neutropenia/ thrombocytopenia/ death due to toxicity/ nausea and vomiting/ diarrhoea/ alopecia/ infection/ asthenia/ pain/ anorexia

Table A3.2 (cont.): Results of the literature search. Treatment option: neoadjuvant chemotherapy with resective intent (A_1)

Article	**Type of study**	**Main treatment/ comparator (if any)**	**Outcome(s) reported**
Katakami, N., H. Tada, and T. Mitsudomi, *A phase 3 study of induction treatment with concurrent chemoradiotherapy versus chemotherapy before surgery in patients with pathologically confirmed N2 stage IIIA nonsmall cell lung cancer (WJTOG9903).* Cancer, 2012. **118**: p. 6126-6135.	Clinical trial	Main treatment: Neoadjuvant chemo-radiotherapy (carboplatin + docetaxel) and surgery Comparator: Neoadjuvant chemotherapy (carboplatin + docetaxel) and surgery	Tumour response/ survival/ disease progression/ nausea/ vomiting/ fever/ dyspnoea/ infection/ peripheral neuropathy/ allergic reaction/ dysphagia/ leukopenia/ neutropenia/ anemia/ thrombocytopenia/ increased transaminase/ increased creatinine
Kolek, V., I. Grygarkova, and M. Hajduch, *Long term follow-up of neoadjuvant-adjuvant combination treatments of IIIA stage non-small cell lung cancer: results of neoadjuvant carboplatin/ vinorelbine and carboplatin/ paclitaxel regimens combined with selective adjuvant chemotherapy according to in vitro chemo-resistance test.* Biomedical Papers, 2008. **152**: p. 259-266.	Clinical trial	Main treatment: Neoadjuvant chemotherapy (carboplatin + vinorelbine) with surgery and post-surgical chemotherapy Comparator: Neoadjuvant chemotherapy (carboplatin + paclitaxel) with surgery and post-surgical chemotherapy	Downstaging/ complete resection/ stable disease/ disease progression/ survival/ overall toxicity/ thrombocytopenia/ myalgia/ arthralgia/ anorexia
Nagai, K., R. Tsuchiya, and T. Mori, *A randomized trial comparing induction chemotherapy followed by surgery with surgery alone for patients with stage IIIA N2 non–small cell lung cancer (JCOG 9209).* The Journal of Thoracic and Cardiovascular Surgery, 2003. **125**: p. 254-260.	Clinical trial	Main treatment: Neoadjuvant chemotherapy (cisplatin + vindesine) with surgery Comparator: Surgery alone	Tumour response/ stable disease/ progressive disease/ survival/ leukocytopenia/ anaemia/ vomiting

From the last column of Table A3.2, the evidence from the literature can be directly used to inform the calculation of the scores of neoadjuvant chemotherapy with resective intent on only a subset of the bottom-level criteria of the decision hierarchy shown in Figure A3.1. Specifically, on the following eleven criteria:

- Life expectancy ($C_2(b)$)

- Asthenia ($C_5(b)$)

- Treatment-related dyspnoea due to pneumonitis or pulmonary fibrosis ($C_{11}(b)$)

- Treatment-related dysphagia due to oesophagitis ($C_{12}(b)$)

- Treatment-related infection due to immunodeficiency ($C_{13}(b)$)

- Treatment-related diarrhoea ($C_{14}(b)$)

- Treatment-related vomiting ($C_{15}(b)$)

- Treatment-related alopecia ($C_{16}(b)$)

- Treatment-related paraesthesia ($C_{17}(b)$)

- Treatment-related anorexia ($C_{19}(b)$)

- Hospital inpatient stays ($C_{21}(b)$)

From the above results, it is likely that the evidence available from the published literature cannot inform the calculation of the scores of this treatment option on the remaining fourteen criteria. Other sources of evidence, in particular clinical expert opinion, are required to calculate these scores.

Table A3.3 presents the outcomes of concurrent chemotherapy (A_2) for Stage $IIIA_3$ NSCLC patients resulting from the five articles selected for review. The table provides, for each article, a brief description of the type of study, the main treatment and comparators described, and the outcomes reported.

Table A3.3: Results of the literature search. Treatment option: concurrent chemo-radiotherapy (A_2)

Article	**Type of study**	**Main treatment/ comparator (if any)**	**Outcome(s) reported**
Auperin, A., et al., Concomitant radio-chemotherapy based on platin compounds in patients with locally advanced non-small cell lung cancer (NSCLC): a meta-analysis of individual data from 1764 patients. Annals of Oncology, 2006. 17(3): p. 473-483	Meta-analysis	Main treatment: radiotherapy with concomitant platin-based chemotherapy Comparator: radiotherapy alone	Survival/ event-free survival
Govindan, R., J. Bogard, and T. Stinchcombe, Randomized phase II study of pemetrexed, carboplatin, and thoracic radiation with or without cetuximab in patients with locally advanced unresectable non–small-cell lung cancer: cancer and leukemia group B trial 30407. Journal of Clinical Oncology, 2011. 29: p. 3120-3125.	Phase II clinical trial	Main treatment: chemotherapy (carboplatin + paclitaxel + cetuximab) with concurrent radiotherapy Main treatment: chemotherapy (carboplatin + paclitaxel + cetuximab) with concurrent radiotherapy	Tumour response/ survival/ anaemia/ neutropenia/ febrile neutropenia/ thrombocytopenia/ dehydration/ dysphagia/ dyspnoea/ esophagitis/ fatigue/ hypokalemia/ nausea and vomiting/ pneumonitis/ rash
Gridelli, C., C. Langer, and P. Maione, Lung cancer in the elderly. Journal of Clinical Oncology, 2007. 25(1898-1907).	Literature review	Range of treatments described: radiotherapy/ concurrent chemo-radiotherapy/ sequential chemoradiotherapy	Survival, quality-adjusted survival/ toxicity

Table A3.3 (cont.). Results of the literature search. Treatment option: concurrent chemo-radiotherapy (A_2)

Article	**Type of study**	**Main treatment/ comparator (if any)**	**Outcome(s) reported**
Hirose, T., Y. Mizutani, and T. Ohmori, The combination of cisplatin and vinorelbine with concurrent thoracic radiation therapy for locally advanced stage IIIA or IIIB non-small-cell lung cancer. Cancer Chemotherapy and Pharmacology, 2006. 58: p. 361-367.	Observational study	Main treatment: chemotherapy (cisplatin + vinorelbine) with concurrent radiotherapy Comparator: N/A	Tumour response/ survival/ leukopenia/ neutropenia/ thrombocytopenia/ anaemia/ nausea/ vomiting/ diarrhoea/ infection/ esophagitis/ pneumonitis/ gastric ulcer/ elevation of transaminase/ elevation of creatinine/ neurological peripheral symptoms
Uitterhoeve, A., M. Koolen, and R. van Os, Accelerated high-dose radiotherapy alone or combined with either concomitant or sequential chemotherapy; treatments of choice in patients with non-small cell lung cancer. Radiation Oncology, 2007. 2: p. 27-36.	Retrospective stud	Main treatment: chemotherapy (cisplatin) with concurrent radiotherapy Comparator 1: chemotherapy (cisplatin + gemcitabine) with sequential radiotherapy Comparator 2: only radiotherapy	Survival/ pulmonary toxicity/ oesophageal toxicity/ cardiac toxicity/ neuropathy

From the last column of Table A3.3, the evidence from the literature can be directly used to inform the calculation of the scores of concurrent chemo-radiotherapy on an even smaller subset of criteria as was the case for neoadjuvant chemotherapy with resective intent. Specifically, on the following seven criteria:

- Life expectancy $(C_2(b))$

- Treatment-related dyspnoea due to pneumonitis or pulmonary fibrosis $(C_{11}(b))$

- Treatment-related dysphagia due to oesophagitis $(C_{12}(b))$

- Treatment-related infection due to immunodeficiency ($C_{13}(b)$)

- Treatment-related diarrhoea ($C_{14}(b)$)

- Treatment-related vomiting ($C_{15}(b)$)

- Treatment-related paraesthesia ($C_{17}(b)$)

As before, it is likely that the evidence available from the published literature cannot inform the calculation of the scores of this treatment option on the remaining fourteen criteria. Other sources of evidence, in particular clinical expert opinion, are required to calculate these scores.

No articles were found to inform the calculation of the scores of the option best supportive care (A_3) on any of the criteria of the decision hierarchy.

To summarise the results of this literature review, it seems likely that the calculation of the scores of the three treatment options - neoadjuvant chemotherapy with resective intent (A_1), concurrent chemo-radiotherapy (A_2) and best supportive care (A_3) on the majority of the bottom-level criteria of the decision hierarchy of the *Expert Choice*-based and the *ALEL*-based PDAs will generally rely on clinical expert opinion.

Appendix 4: *Expert Choice* Evidence-Generation Questionnaire (EC-EGQ), clinicians' judgments about the performance levels of the options on the bottom-level criteria of the hierarchy and single-criterion scores of the options on these criteria for hospital H1

Table A4.1: Expert Choice Evidence Generation Questionnaire (EC-EGC)

Introduction. Welcome. Jose is a sixty-nine year old male recently diagnosed with non-small cell lung cancer stage T2N2M0 ($IIIA_3$). The tumour is in the lower left lobe and the mediastinal involvement is limited to one paratracheal lymph node. Jose has mild chronic obstructive pulmonary disease (COPD) and three years ago he suffered a hear infarction that was treated with a stent. The objective of this questionnaire is to measure the likely impact of the three treatment options available at hospital H1 for this patient on a number of criteria that you have considered of importance to him. The three options are: Option 1: neoadjuvant chemotherapy with resective intent. Option 2: concurrent chemo-radiotherapy Option 3: best supportive care Please answer the following questions. Many thanks in advance for your time and your answers
Question 1. Of 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3), how many of these patients will get "cured"? Cure is defined as absence of tumour activity after 5 years of the start of treatment
Suppose that Jose can be in one of the three states of tumour response: 1) "no progression" (the cancerous lesions do not extend and may even respond partially to the treatment), 2) "progression" (the cancerous lesions extend without question), and 3) "death". **Question 2.** **2.1.** Imagine 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3). **2.1.1.** How many of these patients will be in state "no progression" 6 months after starting treatment? **2.1.2.** How many of these patients will be in state "progression" 6 months after starting treatment? **2.1.3.** How many of these patients will be in state "death" 6 months after starting treatment? **2.2.** Suppose now that Jose, 6 months after starting treatment with (option 1/ Option 2/ Option 3) is in state "no progression". Imagine 100 patients identical to Jose. **2.2.1.** How many of these patients will be in state "no progression" 12 months after starting treatment? **2.2.2.** How many of these patients will be in state "progression" 12 months after starting treatment? **2.2.3.** How many of these patients will be in state "death" 12 months after starting treatment? **2.3.** Suppose now that Jose, 6 months after starting treatment with (option 1/ Option 2/ Option 3) is in state "progression". Imagine 100 patients identical to Jose. **2.3.1.** How many of these patients will still be in state "progression" 12 months after starting treatment? **2.3.2.** How many of these patients will be in state "death" 12 months after starting treatment?

Table A4.1 (cont.): Expert Choice Evidence Generation Questionnaire (EC-EGC)

Question 3. This question is designed to measure Jose's quality of life in the medium term (two years after starting treatment) under each of the three treatment options. To measure this quality of life, it is important to assume that Jose, at two years after starting treatment, is not suffering any of the adverse effects associated with the treatments. In particular, that Jose is not suffering the adverse effects typically associated with chemotherapy or radiotherapy.

3.1. Disease-related pain.
3.1.1. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 2 (or Option 2 over Option 1) in terms of disease related pain two years after Jose starts treatment
3.1.2. Indicate, on a scale between 1 and 9, how much better is Option 2 over Option 3 (or Option 3 over Option 2) in terms of disease related pain two years after Jose starts treatment
3.1.3. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 3 (or Option 3 over Option 1) in terms of disease related pain two years after Jose starts treatment

3.2. Disease-related dyspnoea.
3.2.1. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 2 (or Option 2 over Option 1) in terms of disease related dyspnoea two years after Jose starts treatment
3.2.2. Indicate, on a scale between 1 and 9, how much better is Option 2 over Option 3 (or Option 3 over Option 2) in terms of disease related dyspnoea two years after Jose starts treatment
3.2.3. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 3 (or Option 3 over Option 1) in terms of disease related dyspnoea two years after Jose starts treatment

3.3. Disease-related asthenia.
3.3.1. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 2 (or Option 2 over Option 1) in terms of disease related asthenia two years after Jose starts treatment
3.3.2. Indicate, on a scale between 1 and 9, how much better is Option 2 over Option 3 (or Option 3 over Option 2) in terms of disease related asthenia two years after Jose starts treatment
3.3.3. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 3 (or Option 3 over Option 1) in terms of disease related asthenia two years after Jose starts treatment

3.4. Disease-related emotional problems.
3.4.1. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 2 (or Option 2 over Option 1) in terms of disease related emotional problems two years after Jose starts treatment
3.4.2. Indicate, on a scale between 1 and 9, how much better is Option 2 over Option 3 (or Option 3 over Option 2) in terms of disease related emotional problems two years after Jose starts treatment
3.4.3. Indicate, on a scale between 1 and 9, how much better is Option 1 over Option 3 (or Option 3 over Option 1) in terms of disease related emotional problems two years after Jose starts treatment

Question 4. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to take care of themselves without help from others?

Question 5. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to work a standard working week (i.e. 40 hours)?

Table A4.1 (cont.): Expert Choice Evidence Generation Questionnaire (EC-EGC)

Question 6. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to work a standard working week (i.e. 40 hours)?
Question 7. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). In how many of these patients will the disease interfere from moderately to extremely in their family life and in their family and social relations?
Question 8. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). In how many of these patients will the disease cause moderate to severe financial difficulties?
Question 9. This question is designed to measure the impact of the treatment-related adverse effects on Jose with each of the three treatment options from the start of treatment until death. Out of 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3): **9.1.** How many will suffer grade 2,3, or 4 dyspnoea as a consequence of treatment-related pneumonitis and/or pulmonary fibrosis? **9.2.** How many will suffer grade 2,3, or 4 dysphagia as a consequence of treatment-related esophagitis? **9.3.** How many will suffer grade 2,3, or 4 infection as a consequence of immunodeficiency? **9.4.** How many will suffer grade 2,3, or 4 treatment-related diarrhoea? **9.5.** How many will suffer grade 2,3, or 4 treatment-related vomiting? **9.6.** How many will suffer grade 2 alopecia? **9.7.** How many will suffer grade 2 or grade 3 paraesthesia? **9.8.** How many will suffer grade 2 or grade 3 fatigue? **9.9.** How many will suffer grade 2, 3, or 4 anorexia?
Question 10. This question is designed to measure the quality of the health care experience for Jose, from the start of treatment with (Option 1/ Option 2/ Option 3) until two years after the start of treatment. **10.1.** On average, what is the total number of visits that Jose will make to any outpatient health service during this period? **10.2.** On average, what is the total number of days spent by Jose in the hospital due to a cancer-related hospitalisation during this period? **10.3.** On average, what is the total number of days over the optimal calendar required to continue treating Jose due to waiting lists during this period? **10.4.** Will Jose be treated by the same clinician or team of clinicians during this period? **10.5.** Will Jose be treated by his clinician or team of clinicians in a caring and considerate fashion during this period?

Table A4.2. Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (Expert Choice-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$ or Matrix of comparative consequence judgments $CM(C_k(b))$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Cure $C_1(b)$	$x_{i,1}$	0.25	0.10	0.05	0.20	0.10	0.00	0.40	0.27	0.00
Life Expectancy $C_2(b)$	$x_{i,2}$	5.15	5.05	1.75	2.43	1.34	0.55	6.49	3.13	0.73
Disease-related pain $C_3(b)$	$CM(C_3(b))$	$\begin{pmatrix} 1 & 3 & 5 \\ 1/4 & 1 & 4 \\ 1/5 & 1/3 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 4 & 3 \\ 1/4 & 1 & 1 \\ 1/3 & 1 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 3 & 7 \\ 1/3 & 1 & 5 \\ 1/7 & 1/5 & 1 \end{pmatrix}$		
Disease-related dyspnoea $C_4(b)$	$CM(C_4(b))$	$\begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 4 \\ 1/5 & 1/4 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 4 & 5 \\ 1/4 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 1 & 8 \\ 1 & 1 & 7 \\ 1/8 & 1/7 & 1 \end{pmatrix}$		
Disease-related asthenia $C_5(b)$	$CM(C_5(b))$	$\begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 4 \\ 1/5 & 1/4 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 4 & 3 \\ 1/4 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 1 & 9 \\ 1 & 1 & 9 \\ 1/9 & 1/9 & 1 \end{pmatrix}$		
Disease-related emotional problems $C_6(b)$	$CM(C_6(b))$	$\begin{pmatrix} 1 & 5 & 5 \\ 1/5 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 4 & 5 \\ 1/4 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix}$			$\begin{pmatrix} 1 & 3 & 8 \\ 1/3 & 1 & 6 \\ 1/8 & 1/6 & 1 \end{pmatrix}$		

Table A4.2 (cont). Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (Expert Choice-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$ or Matrix of comparative consequence judgments $CM(C_k(b))$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Self-care $C_7(b)$	$x_{i,7}$	0.90	0.80	0.30	0.90	0.80	0.10	0.85	0.60	0.10
Work a normal week $C_8(b)$	$x_{i,8}$	0.25	0.05	0.00	0.80	0.60	0.20	0.75	0.55	0.00
Interference of the disease with family life or with social activities $C_9(b)$	$x_{i,9}$	0.20	0.40	0.90	0.20	0.40	0.80	0.15	0.30	0.99
Disease-related financial burden in the medium term* $C_{10}(b)$	$x_{i,10}$	0.60	0.90	0.95	0.20	0.40	0.80	0.20	0.35	0.99
Treatment-related dyspnoea as a consequence of pneumonitis or pulmonary fibrosis $C_{11}(b)$	$x_{i,11}$	0.20	0.50	0.00	0.10	0.40	0.00	0.02	0.04	0.00

Table A4.2(cont). Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (Expert Choice-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related dysphagia as a consequence of oesophagitis $C_{12}(b)$	$x_{i,12}$	0.10	0.20	0.00	0.15	0.40	0.00	0.02	0.15	0.00
Treatment-related infection due to immuno-deficiency $C_{13}(b)$	$x_{i,13}$	0.20	0.20	0.10	0.20	0.30	0.00	0.02	0.20	0.00
Treatment-related diarrhoea $C_{14}(b)$	$x_{i,14}$	0.20	0.20	0.10	0.10	0.10	0.05	0.20	0.30	0.05
Treatment-related vomiting $C_{15}(b)$	$x_{i,15}$	0.20	0.20	0.10	0.15	0.10	0.05	0.15	0.60	0.20
Treatment-related alopecia $C_{16}(b)$	$x_{i,16}$	0.10	0.10	0.00	0.20	0.20	0.00	0.05	0.60	0.00
Treatment-related paraesthesia $C_{17}(b)$	$x_{i,17}$	0.05	0.05	0.00	0.10	0.10	0.00	0.10	0.50	0.00

Table A4.2 (cont). Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (Expert Choice-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related fatigue $C_{18}(b)$	$x_{i,18}$	0.40	0.50	0.20	0.40	0.60	0.05	0.02	0.30	0.00
Treatment-related anorexia $C_{19}(b)$	$x_{i,19}$	0.40	0.40	0.00	0.60	0.80	0.00	0.02	0.30	0.00
Visits to the health services $C_{20}(b)$	$x_{i,20}$	20	20	10	46	76	20	13	31	4
Hospital in-patient stays $C_{21}(b)$	$x_{i,21}$	30	20	10	15	10	10	14	14	21
Waiting time (due to waiting lists) between interventions $C_{22}(b)$	$x_{i,22}$	10	20	10	45	50	15	30	15	0
Treatment by the same team of clinicians $C_{23}(b)$	$x_{i,23}$	1	1	1	1	1	1	1	1	1
Attentive care $C_{24}(b)$	$x_{i,24}$	1	1	1	1	1	1	1	1	1

Table A4.3. Transition probability matrices for the Stage IIIA$_3$ NSCLC Markov model elicited from the three clinicians in hospital H1 (Expert Choice-based PDA)

Option A_i $(i = 1 \dots 3)$	**Clinician 1**	**Clinician 2**	**Clinician 3**
Neoadjuvant chemotherapy with resective intent (A_1)	$\begin{pmatrix} 0 & 0.80 & 0.15 & 0.05 \\ 0 & 0.80 & 0.15 & 0.05 \\ 0 & 0 & 0.90 & 0.10 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.90 & 0.08 & 0.02 \\ 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0 & 0.50 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.95 & 0.05 & 0 \\ 0 & 0.92 & 0.04 & 0.04 \\ 0 & 0 & 0.90 & 0.10 \end{pmatrix}$
Chemo-radiotherapy (A_2)	$\begin{pmatrix} 0 & 0.70 & 0.25 & 0.05 \\ 0 & 0.80 & 0.15 & 0.05 \\ 0 & 0 & 0.90 & 0.10 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0.50 & 0.30 & 0.20 \\ 0 & 0 & 0.20 & 0.80 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.95 & 0.04 & 0.01 \\ 0 & 0.75 & 0.20 & 0.05 \\ 0 & 0 & 0.60 & 0.40 \end{pmatrix}$
Best supportive care (A_3)	$\begin{pmatrix} 0 & 0.20 & 0.70 & 0.10 \\ 0 & 0.30 & 0.40 & 0.30 \\ 0 & 0 & 0.70 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.10 & 0.45 & 0.45 \\ 0 & 0.02 & 0.18 & 0.80 \\ 0 & 0 & 0.05 & 0.95 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.20 & 0.55 & 0.25 \\ 0 & 0.05 & 0.55 & 0.40 \\ 0 & 0 & 0.10 & 0.90 \end{pmatrix}$

Table A4.4. Levels $x_{i,2}$ $(i = 1 \dots 3)$ of the three options A_i $(i = 1 \dots 3)$ on criterion life expectancy $C_2(b)$, in years (Expert Choice-based PDA)

Option A_i $(i = 1 \dots 3)$	**Clinician 1**	**Clinician 2**	**Clinician 3**
Neoadjuvant chemotherapy with resective intent (A_1)	$x_{1,2} = 5.15$	$x_{1,2} = 2.43$	$x_{1,2} = 6.49$
Chemo-radiotherapy (A_2)	$x_{2,2} = 5.05$	$x_{2,2} = 1.34$	$x_{2,2} = 3.13$
Best supportive care (A_3)	$x_{3,2} = 1.75$	$x_{3,2} = 0.55$	$x_{3,2} = 0.73$

Table A4.5. Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (Expert Choice-based PDA)

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Cure $C_1(b)$	0.33 (1.00)	0.11 (0.33)	0.05 (0.16)	0.69 (1)	0.31 (0.44)	0.00 (0.00)	0.64 (1.00)	0.36 (0.55)	0.00 (0.00)
Life Expectancy $C_2(b)$	1.00 (1.00)	0.98 (0.98)	0.34 (0.34)	1.00 (1.00)	0.55 (0.55)	0.23 (0.23)	1.00 (1.00)	0.48 (0.48)	0.11 (0.11)
Disease-related pain $C_3(b)$	0.67 (1.00)	0.23 (0.34)	0.10 (0.15)	0.63 (1.00)	0.17 (0.28)	0.19 (0.30)	0.65 (1.00)	0.28 (0.43)	0.07 (0.11)
Disease-related dyspnoea $C_4(b)$	0.63 (1.00)	0.28 (0.44)	0.09 (0.14)	0.67 (1.00)	0.23 (0.32)	0.10 (0.15)	0.48 (1.00)	0.46 (0.96)	0.06 (0.13)
Disease-related asthenia $C_5(b)$	0.63 (1.00)	0.28 (0.44)	0.09 (0.14)	0.63 (1.00)	0.22 (0.35)	0.15 (0.24)	0.47 (1.00)	0.47 (1.00)	0.05 (0.11)
Disease-related emotional problems $C_6(b)$	0.71 (1.00)	0.18 (0.25)	0.11 (0.16)	0.67 (1.00)	0.23 (0.32)	0.10 (0.15)	0.65 (1.00)	0.29 (0.44)	0.06 (0.10)
Self-care $C_7(b)$	0.67 (1.00)	0.30 (0.45)	0.03 (0.05)	0.69 (1.00)	0.31 (0.44)	0.01 (0.01)	0.78 (1.00)	0.21 (0.26)	0.02 (0.02)
Work a normal week $C_8(b)$	0.86 (1.00)	0.14 (0.16)	0.00 (0.00)	0.70 (1.00)	0.26 (0.38)	0.04 (0.06)	0.71 (1.00)	0.29 (0.41)	0.00 (0.00)
Interference of the disease with family life or with social activities $C_9(b)$	0.71 (1.00)	0.27 (0.38)	0.02 (0.03)	0.70 (1.00)	0.26 (0.38)	0.04 (0.06)	0.71 (1.00)	0.29 (0.41)	0.00 (0.00)
Disease-related financial burden in the medium term $C_{10}(b)$	0.80 (1.00)	0.13 (0.16)	0.06 (0.08)	0.70 (1.00)	0.26 (0.38)	0.04 (0.06)	0.68 (1.00)	0.32 (0.46)	0.00 (0.00)

*Table A4.5 (cont): Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (Expert Choice-*based *PDA)*

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related dyspnoea as a consequence of pneumonitis or pulmonary fibrosis $C_{11}(b)$	0.04 (0.04)	0.01 (0.01)	0.95 (1.00)	0.01 (0.09)	0.08 (0.02)	0.90 (1.00)	0.28 (0.49)	0.14 (0.24)	0.58 (1.00)
Treatment-related dysphagia as a consequence of oesophagitis $C_{12}(b)$	0.08 (0.09)	0.04 (0.04)	0.88 (1.00)	0.05 (0.06)	0.01 (0.02)	0.93 (1.00)	0.32 (0.49)	0.04 (0.06)	0.64 (1.00)
Treatment-related infection due to immuno-deficiency $C_{13}(b)$	0.04 (0.04)	0.04 (0.04)	0.92 (1.00)	0.02 (0.02)	0.04 (0.04)	0.94 (1.00)	0.04 (0.04)	0.02 (0.02)	0.94 (1.00)
Treatment-related diarrhoea $C_{14}(b)$	0.24 (0.46)	0.24 (0.46)	0.54 (1.00)	0.24 (0.46)	0.24 (0.46)	0.54 (1.00)	0.16 (0.21)	0.09 (0.12)	0.75 (1.00)
Treatment-related vomiting $C_{15}(b)$	0.24 (0.46)	0.24 (0.46)	0.53 (1.00)	0.17 (0.30)	0.27 (0.47)	0.56 (1.00)	0.02 (0.02)	0.00 (0.00)	0.98 (1.00)
Treatment-related alopecia $C_{16}(b)$	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)
Treatment-related paraesthesia $C_{17}(b)$	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)
Treatment-related fatigue $C_{18}(b)$	0.23 (0.37)	0.15 (0.24)	0.62 (1.00)	0.07 (0.08)	0.03 (0.04)	0.90 (1.00)	0.33 (0.50)	0.02 (0.02)	0.66 (1.00)

*Table A4.5 (cont): Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (Expert Choice-*based *PDA)*

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related anorexia $C_{19}(b)$	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.33 (0.50)	0.02 (0.02)	0.66 (1.00)
Visits to the health services $C_{20}(b)$	0.25 (0.50)	0.25 (0.50)	0.50 (1.00)	0.26 (0.43)	0.15 (0.26)	0.59 (1.00)	0.21 (0.31)	0.09 (0.13)	0.70 (1.00)
Hospital in-patient stays $C_{21}(b)$	0.18 (0.33)	0.27 (0.49)	0.55 (1.00)	0.25 (0.67)	0.38 (1.00)	0.38 (1.00)	0.37 (1.00)	0.38 (1.00)	0.25 (0.67)
Waiting time (due to waiting lists) between interventions $C_{22}(b)$	0.40 (1.00)	0.20 (0.50)	0.40 (1.00)	0.20 (0.33)	0.18 (0.30)	0.62 (1.00)	0.00 (0.00)	0.01 (0.01)	0.99 (1.00)
Treatment by the same team of clinicians $C_{23}(b)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Attentive care $C_{24}(b)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)

Appendix 5: *ALEL* Evidence-Generation Questionnaire (*ALEL*-EGQ), clinicians' judgments about the performance levels of the options on the bottom-level criteria of the hierarchy and single-criterion scores of the options on these criteria for *ALEL*-based PDA

Table A5.1. ALEL-based evidence generation questionnaire (ALEL-EGC)

Introduction. Welcome. Jose is a sixty-nine year old male recently diagnosed with non-small cell lung cancer stage T2N2M0 ($IIIA_3$). The tumour is in the lower left lobe and the mediastinal involvement is limited to one paratracheal lymph node. Jose has mild chronic obstructive pulmonary disease (COPD) and three years ago he suffered a hear infarction that was treated with a stent. The objective of this questionnaire is to measure the likely impact of the three treatment options available at hospital H1 for this patient on a number of criteria that you have considered of importance to him. The three options are: Option 1: neoadjuvant chemotherapy with resective intent. Option 2: concurrent chemo-radiotherapy Option 3: best supportive care Please answer the following questions. Many thanks in advance for your time and your answers
Question 1. Of 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3), how many of these patients will get "cured"? Cure is defined as absence of tumour activity after 5 years of the start of treatment
Suppose that Jose can be in one of the three states of tumour response: 1) "no progression" (the cancerous lesions do not extend and may even respond partially to the treatment), 2) "progression" (the cancerous lesions extend without question), and 3) "death". **Question 2.** **2.1.** Imagine 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3). **2.1.1.** How many of these patients will be in state "no progression" 6 months after starting treatment? **2.1.2.** How many of these patients will be in state "progression" 6 months after starting treatment? **2.1.3.** How many of these patients will be in state "death" 6 months after starting treatment? **2.2.** Suppose now that Jose, 6 months after starting treatment with (option 1/ Option 2/ Option 3) is in state "no progression". Imagine 100 patients identical to Jose. **2.2.1.** How many of these patients will be in state "no progression" 12 months after starting treatment? **2.2.2.** How many of these patients will be in state "progression" 12 months after starting treatment? **2.2.3.** How many of these patients will be in state "death" 12 months after starting treatment? **2.3.** Suppose now that Jose, 6 months after starting treatment with (option 1/ Option 2/ Option 3) is in state "progression". Imagine 100 patients identical to Jose. **2.3.1.** How many of these patients will still be in state "progression" 12 months after starting treatment? **2.3.2.** How many of these patients will be in state "death" 12 months after starting treatment?

Table A5.1 (cont.). ALEL-based evidence generation questionnaire (ALEL-EGC)

Question 3. This question is designed to measure Jose's quality of life in the medium term (two years after starting treatment) under each of the three treatment options. To measure this quality of life, it is important to assume that Jose, at two years after starting treatment, is not suffering any of the adverse effects associated with the treatments. In particular, that Jose is not suffering the adverse effects typically associated with chemotherapy or radiotherapy. **3.1.** Indicate, on a scale between 1 and 9 where 1 is "no pain" and 9 is "extreme pain", the intensity of disease-related pain that Jose will experience two years after the start of treatment with (Option 1/ Option 2/ Option 3) **3.2.** Indicate, on a scale between 1 and 9 where 1 is "no dyspnoea" and 9 is "extreme dyspnoea", the intensity of disease-related dyspnoea that Jose will experience two years after the start of treatment with (Option 1/ Option 2/ Option 3) **3.3.** Indicate, on a scale between 1 and 9 where 1 is "no asthenia" and 9 is "extreme asthenia", the intensity of disease-related asthenia that Jose will experience two years after the start of treatment with (Option 1/ Option 2/ Option 3) **3.4.** Indicate, on a scale between 1 and 9 where 1 is "no emotional problems" and 9 is "extreme emotional problems", the intensity of disease-related emotional problems that Jose will experience two years after the start of treatment with (Option 1/ Option 2/ Option 3)
Question 4. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to take care of themselves without help from others?
Question 5. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to work a standard working week (i.e. 40 hours)?
Question 6. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). How many of these patients will be able to work a standard working week (i.e. 40 hours)?
Question 7. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). In how many of these patients will the disease interfere from moderately to extremely in their family life and in their family and social relations?
Question 8. Imagine 100 patients identical to Jose who are alive two years after starting treatment with (Option 1/ Option 2/ Option 3). In how many of these patients will the disease cause moderate to severe financial difficulties?

Table A5.1 (cont.): ALEL-based evidence generation questionnaire (ALEL-EGC)

Question 9. This question is designed to measure the impact of the treatment-related adverse effects on Jose with each of the three treatment options from the start of treatment until death.

Out of 100 patients identical to Jose who start treatment with (Option 1/ Option 2/ Option 3):
9.1. How many will suffer grade 2,3, or 4 dyspnoea as a consequence of treatment-related pneumonitis and/or pulmonary fibrosis?
9.2. How many will suffer grade 2,3, or 4 dysphagia as a consequence of treatment-related esophagitis?
9.3. How many will suffer grade 2,3, or 4 infection as a consequence of immunodeficiency?
9.4. How many will suffer grade 2,3, or 4 treatment-related diarrhoea?
9.5. How many will suffer grade 2,3, or 4 treatment-related vomiting?
9.6. How many will suffer grade 2 alopecia?
9.7. How many will suffer grade 2 or grade 3 paraesthesia?
9.8. How many will suffer grade 2 or grade 3 fatigue?
9.9. How many will suffer grade 2, 3, or 4 anorexia?

Question 10. This question is designed to measure the quality of the health care experience for Jose, from the start of treatment with (Option 1/ Option 2/ Option 3) until two years after the start of treatment.

10.1. On average, what is the total number of visits that Jose will make to any outpatient health service during this period?
10.2. On average, what is the total number of days spent by Jose in the hospital due to a cancer-related hospitalisation during this period?
10.3. On average, what is the total number of days over the optimal calendar required to continue treating Jose due to waiting lists during this period?
10.4. Will Jose be treated by the same clinician or team of clinicians during this period?
10.5. Will Jose be treated by his clinician or team of clinicians in a caring and considerate fashion during this period?

Table A5.2. Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (ALEL-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Cure $C_1(b)$	$x_{i,1}$	0.30	0.10	0.03	0.40	0.25	0.00	0.25	0.15	0.00
Life Expectancy $C_2(b)$	$x_{i,2}$	1.85	1.19	1.12	2.08	1.28	0.70	2.01	1.23	0.38
Disease-related pain $C_3(b)$	$x_{i,3}$	3	4	4	5	6	7	2	3	4
Disease-related dyspnoea $C_4(b)$	$x_{i,4}$	4	6	4	5	6	7	2	5	7
Disease-related asthenia $C_5(b)$	$x_{i,5}$	4	6	8	3	4	6	2	4	7
Disease-related emotional problems $C_6(b)$	$x_{i,6}$	4	6	8	3	3	6	2	4	6
Self-care $C_7(b)$	$x_{i,7}$	0.60	0.35	0.05	0.70	0.50	0.25	0.80	0.60	0.10
Work a normal week $C_8(b)$	$x_{i,8}$	0.40	0.20	0.05	0.30	0.20	0.10	0.80	0.70	0.10

Table A5.2 (cont.). Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (ALEL-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Interference of the disease with family life or with social activities $C_9(b)$	$x_{i,9}$	0.40	0.65	0.05	0.50	0.60	0.80	0.80	0.70	0.50
Disease-related financial burden in the medium term* $C_{10}(b)$	$x_{i,10}$	0.10	0.20	0.10	0.50	0.60	0.80	0.10	0.10	0.20
Treatment-related dyspnoea as a consequence of pneumonitis or pulmonary fibrosis $C_{11}(b)$	$x_{i,11}$	0.05	0.20	0.00	0.50	0.70	0.00	0.15	0.30	0.00
Treatment-related dysphagia as a consequence of oesophagitis $C_{12}(b)$	$x_{i,12}$	0.10	0.20	0.00	0.30	0.60	0.00	0.10	0.20	0.00
Treatment-related infection due to immuno-deficiency $C_{13}(b)$	$x_{i,13}$	0.30	0.50	0.00	0.10	0.20	0.00	0.15	0.30	0.00

Table A5.2 (cont.): Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (ALEL-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related diarrhoea $C_{14}(b)$	$x_{i,14}$	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.05
Treatment-related vomiting $C_{15}(b)$	$x_{i,15}$	0.05	0.30	0.10	0.50	0.60	0.10	0.20	0.30	0.30
Treatment-related alopecia $C_{16}(b)$	$x_{i,16}$	0.90	0.99	0.00	0.20	0.20	0.00	0.80	0.80	0.00
Treatment-related paraesthesia $C_{17}(b)$	$x_{i,17}$	0.10	0.50	0.00	0.20	0.20	0.00	0.70	0.70	0.00
Treatment-related fatigue $C_{18}(b)$	$x_{i,18}$	0.20	0.40	0.60	0.40	0.60	0.15	0.15	0.25	0.50
Treatment-related anorexia $C_{19}(b)$	$x_{i,19}$	0.20	0.50	0.00	0.15	0.40	0.00	0.30	0.50	0.00
Visits to the health services $C_{20}(b)$	$x_{i,20}$	8	20	4	20	30	10	15	25	5

Table A5.2 (cont.). Levels of the three options A_i $(i = 1 \ldots 3)$ on the bottom-level criteria of the full final hierarchy for each of the three clinicians (ALEL-based PDA)

Criterion $C_k(b)$	**Level $x_{i,k}$**	**Clinician 1**			**Clinician 2**			**Clinician 3**		
		A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Hospital in-patient stays $C_{21}(b)$	$x_{i,21}$	10	15	4	20	20	30	10	15	5
Waiting time (due to waiting lists) between interventions $C_{22}(b)$	$x_{i,22}$	30	20	5	15	30	10	5	10	1
Treatment by the same team of clinicians $C_{23}(b)$	$x_{i,23}$	1	1	1	1	1	1	1	1	1
Attentive care $C_{24}(b)$	$x_{i,24}$	1	1	1	1	1	1	1	1	1

Table A5.3. Transition probability matrices for the Stage IIIA$_3$ NSCLC Markov model elicited from the three clinicians in hospital H2 (ALEL-based PDA)

Option A_i $(i = 1 \ldots 3)$	**Clinician 1**	**Clinician 2**	**Clinician 3**
Neoadjuvant chemotherapy with resective intent (A_1)	$\begin{pmatrix} 0 & 0.75 & 0.18 & 0.07 \\ 0 & 0.55 & 0.35 & 0.10 \\ 0 & 0 & 0.50 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.75 & 0.20 & 0.05 \\ 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0 & 0.40 & 0.60 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.80 & 0.15 & 0.05 \\ 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0 & 0.80 & 0.20 \end{pmatrix}$
Chemo-radiotherapy (A_2)	$\begin{pmatrix} 0 & 0.50 & 0.35 & 0.15 \\ 0 & 0.40 & 0.40 & 0.20 \\ 0 & 0 & 0.35 & 0.65 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.60 & 0.30 & 0.10 \\ 0 & 0.50 & 0.25 & 0.25 \\ 0 & 0 & 0.30 & 0.70 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.60 & 0.25 & 0.15 \\ 0 & 0.50 & 0.30 & 0.20 \\ 0 & 0 & 0.20 & 0.80 \end{pmatrix}$
Best supportive care (A_3)	$\begin{pmatrix} 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0.15 & 0.60 & 0.25 \\ 0 & 0 & 0.25 & 0.75 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.30 & 0.30 & 0.40 \\ 0 & 0.30 & 0.30 & 0.40 \\ 0 & 0 & 0.1 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.05 & 0.20 & 0.75 \\ 0 & 0.01 & 0.04 & 0.95 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Table A5.4. Levels $x_{i,2}$ $(i = 1 \ldots 3)$ of the three options A_i $(i = 1 \ldots 3)$ on criterion life expectancy $C_2(b)$, in years (ALEL-based PDA)

Option A_i $(i = 1 \ldots 3)$	**Clinician 1**	**Clinician 2**	**Clinician 3**
Neoadjuvant chemotherapy with resective intent (A_1)	$x_{1,2} = 1.85$	$x_{1,2} = 2.08$	$x_{1,2} = 2.01$
Chemo-radiotherapy (A_2)	$x_{2,2} = 1.19$	$x_{2,2} = 1.24$	$x_{2,2} = 1.23$
Best supportive care (A_3)	$x_{3,2} = 1.12$	$x_{3,2} = 0.70$	$x_{3,2} = 0.38$

Table A5.5. Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (ALEL-based PDA)

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Cure $C_1(b)$	0.30 (1.00)	0.10 (0.33)	0.03 (0.10)	0.40 (1.00)	0.25 (0.63)	0.00 (0.00)	0.25 (1.00)	0.15 (0.6)	0.00 (0.00)
Life Expectancy $C_2(b)$	1.00 (1.00)	0.64 (0.64)	0.61 (0.61)	1.00 (1.00)	0.62 (0.62)	0.34 (0.34)	1.00 (1.00)	0.61 (0.61)	0.19 (0.19)
Disease-related pain $C_3(b)$	0.75 (1.00)	0.63 (0.83)	0.63 (0.83)	0.50 (1.00)	0.38 (0.75)	0.25 (0.50)	0.88 (1.00)	0.75 (0.86)	0.63 (0.71)
Disease-related dyspnoea $C_4(b)$	0.63 (1.00)	0.38 (0.60)	0.63 (1.00)	0.50 (1.00)	0.38 (0.75)	0.25 (0.50)	0.88 (1.00)	0.50 (0.57)	0.25 (0.29)
Disease-related asthenia $C_5(b)$	0.63 (1.00)	0.38 (0.60)	0.13 (0.20)	0.75 (1.00)	0.63 (0.83)	0.38 (0.50)	0.88 (1.00)	0.63 (0.71)	0.25 (0.29)
Disease-related emotional problems $C_6(b)$	0.63 (1.00)	0.38 (0.60)	0.13 (0.20)	0.75 (1.00)	0.75 (1.00)	0.38 (0.50)	0.88 (1.00)	0.63 (0.71)	0.38 (0.43)
Self-care $C_7(b)$	0.60 (1.00)	0.35 (0.58)	0.05 (0.08)	0.70 (1.00)	0.50 (0.71)	0.25 (0.36)	0.80 (1.00)	0.60 (0.75)	0.10 (0.13)
Work a normal week $C_8(b)$	0.40 (1.00)	0.20 (0.50)	0.05 (0.13)	0.30 (1.00)	0.20 (0.67)	0.10 (0.33)	0.80 (1.00)	0.70 (0.88)	0.10 (0.13)
Interference of the disease with family life or with social activities $C_9(b)$	0.60 (0.63)	0.35 (0.37)	0.95 (1.00)	0.50 (1.00)	0.40 (0.80)	0.20 (0.40)	0.71 (1.00)	0.29 (0.41)	0.00 (0.00)
Disease-related financial burden in the medium term $C_{10}(b)$	0.90 (1.00)	0.80 (0.89)	0.90 (1.00)	0.50 (1.00)	0.40 (0.80)	0.20 (0.40)	0.50 (1.00)	0.30 (0.60)	0.20 (0.40)

*Table A5.5 (cont.). Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (ALEL-*based *PDA)*

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related dyspnoea as a consequence of pneumonitis or pulmonary fibrosis $C_{11}(b)$	0.95 (0.95)	0.80 (0.80)	1.00 (1.00)	0.50 (0.50)	0.30 (0.30)	1.00 (1.00)	0.85 (0.85)	0.70 (0.70)	1.00 (1.00)
Treatment-related dysphagia as a consequence of oesophagitis $C_{12}(b)$	0.90 (0.90)	0.80 (0.80)	1.00 (1.00)	0.70 (0.70)	0.40 (0.40)	1.00 (1.00)	0.90 (0.90)	0.80 (0.80)	1.00 (1.00)
Treatment-related infection due to immuno-deficiency $C_{13}(b)$	0.70 (0.70)	0.50 (0.50)	1.00 (1.00)	0.90 (0.90)	0.80 (0.80)	1.00 (1.00)	0.85 (0.85)	0.70 (0.70)	1.00 (1.00)
Treatment-related diarrhoea $C_{14}(b)$	0.90 (1.00)	0.80 (0.89)	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (0.95)	0.90 (0.95)	0.95 (1.00)
Treatment-related vomiting $C_{15}(b)$	0.95 (1.00)	0.70 (0.74)	0.90 (0.95)	0.50 (0.55)	0.40 (0.44)	0.90 (1.00)	0.80 (1.00)	0.70 (0.88)	0.70 (0.88)
Treatment-related alopecia $C_{16}(b)$	0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	0.80 (0.80)	0.80 (0.80)	1.00 (1.00)	0.20 (0.20)	0.20 (0.20)	1.00 (1.00)
Treatment-related paraesthesia $C_{17}(b)$	0.9 (0.9)	0.5 (0.5)	1.00 (1.00)	0.80 (0.80)	0.80 (0.80)	1.00 (1.00)	0.30 (0.30)	0.30 (0.30)	1.00 (1.00)
Treatment-related fatigue $C_{18}(b)$	0.80 (1.00)	0.60 (0.75)	0.40 (0.5)	0.60 (0.71)	0.40 (0.47)	0.85 (1.00)	0.85 (1.00)	0.75 (0.88)	0.50 (0.59)

*Table A5.5 (cont.): Non-idealised scores (idealised scores) of the options $A_i(i = 1 \dots 3)$ on the bottom-level criteria $C_k(b)(k = 1 \dots 24)$ resulting from clinicians' judgments (ALEL-*based *PDA)*

Criterion $C_k(b)$	**Clinician 1**			**Clinician 2**			**Clinician 3**		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
Treatment-related anorexia $C_{19}(b)$	0.80 (0.80)	0.50 (0.5)	1.00 (1.00)	0.85 (0.85)	0.60 (0.60)	1.00 (1.00)	0.70 (0.70)	0.50 (0.50)	1.00 (1.00)
Visits to the health services $C_{20}(b)$	0.50 (0.50)	0.20 (0.20)	1.00 (1.00)	0.50 (0.50)	0.33 (0.33)	1.00 (1.00)	0.07 (0.33)	0.04 (0.20)	0.20 (1)
Hospital in-patient stays $C_{21}(b)$	0.40 (0.40)	0.27 (0.27)	1.00 (1.00)	1.00 (1.00)	0.67 (0.67)	0.67 (0.67)	0.10 (0.5)	0.07 (0.33)	0.20 (1.00)
Waiting time (due to waiting lists) between interventions $C_{22}(b)$	0.17 (0.17)	0.25 (0.25)	1.00 (1.00)	0.67 (0.67)	0.33 (0.33)	1.00 (1.00)	0.20 (1.00)	0.10 (0.59)	0.10 (0.50)
Treatment by the same team of clinicians $C_{23}(b)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Attentive care $C_{24}(b)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)

Appendix 6. Visual representation of the sequence of delivery steps using the *Expert Choice*-based PDA (hospital H1)

Figure A6.1. Communication of the criteria

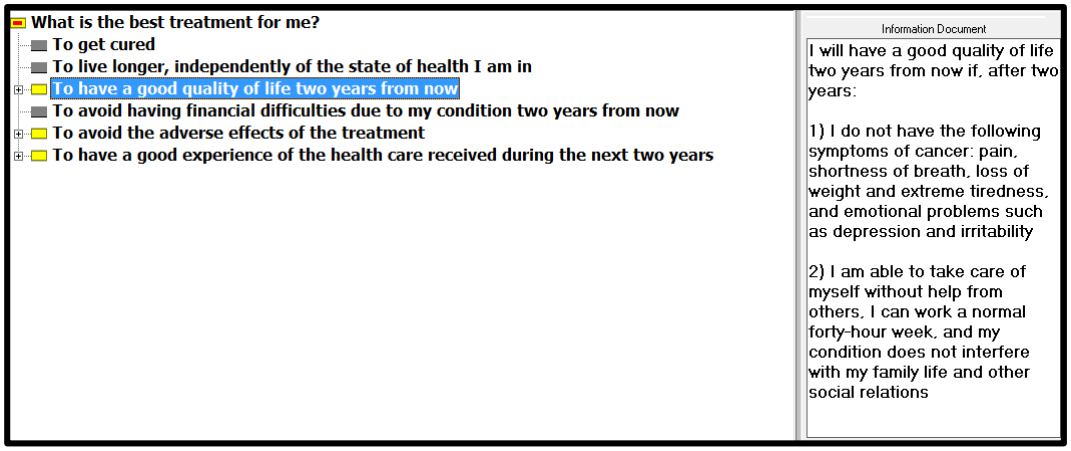


Figure A6.2. Pair-wise comparisons screen

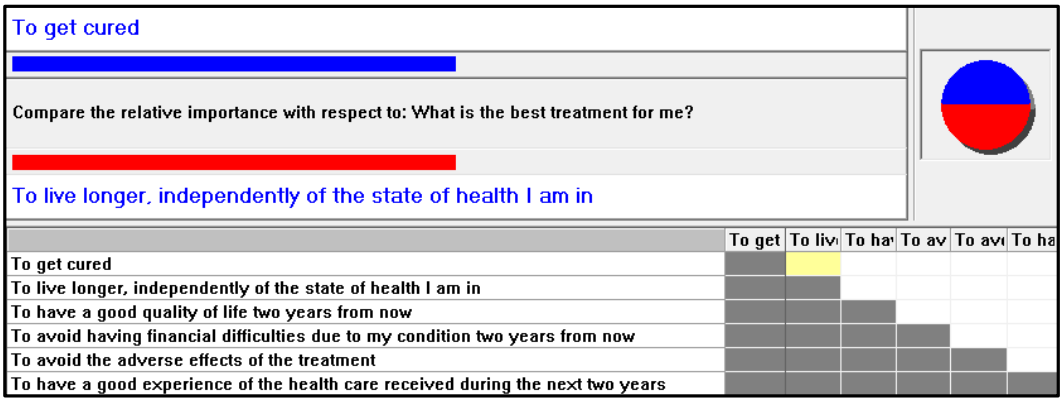


Figure A6.3. Communication of the options

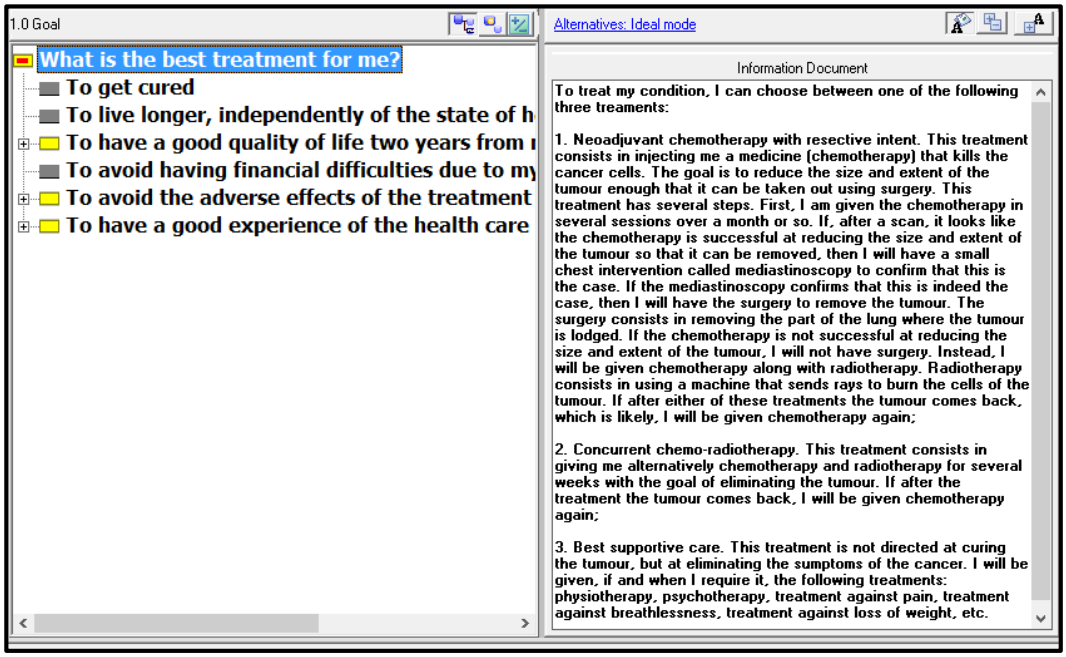


Figure A6.4. Communication of the results of the decision

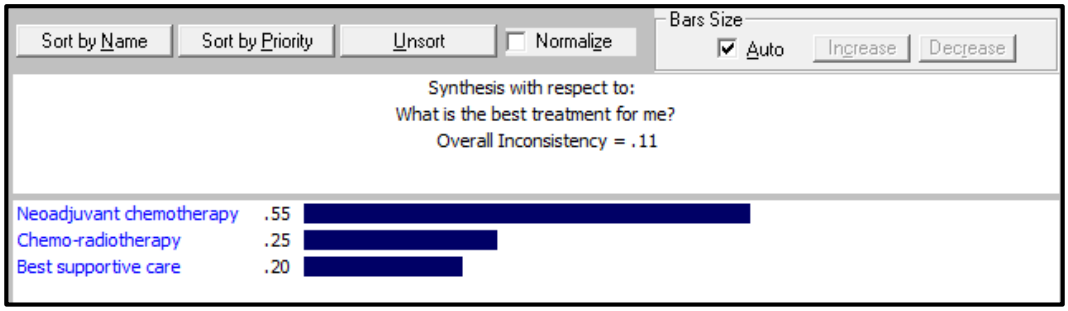


Figure A6.5. Communication of the evidence for the quality of life in the medium term criterion

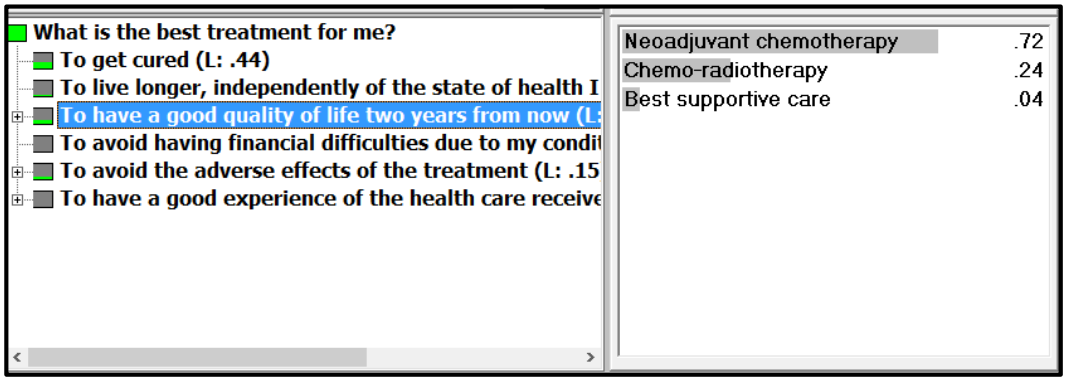
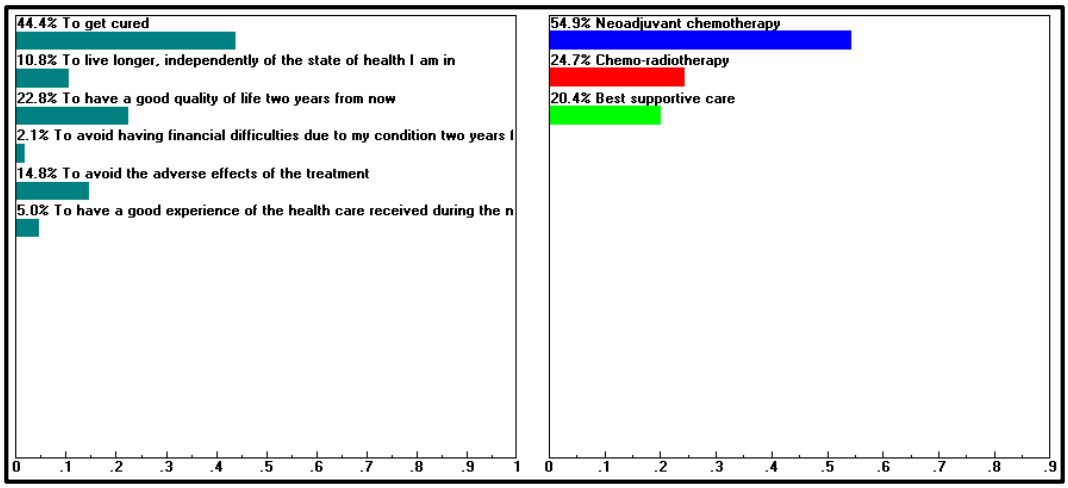
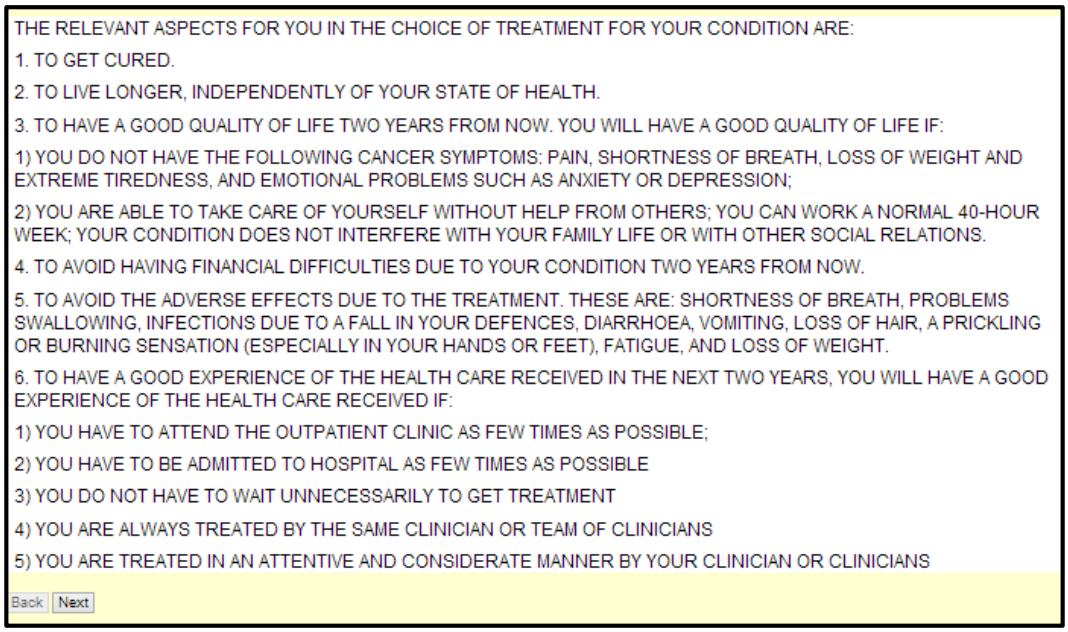


Figure A6.6. Dynamic sensitivity analysis screen



Appendix 7. Visual representation of the sequence of delivery steps using the *ALEL*-based PDA (hospital H2)

Figure A7.1. Communication of the criteria


THE RELEVANT ASPECTS FOR YOU IN THE CHOICE OF TREATMENT FOR YOUR CONDITION ARE:

1. TO GET CURED.

2. TO LIVE LONGER, INDEPENDENTLY OF YOUR STATE OF HEALTH.

3. TO HAVE A GOOD QUALITY OF LIFE TWO YEARS FROM NOW. YOU WILL HAVE A GOOD QUALITY OF LIFE IF:

1) YOU DO NOT HAVE THE FOLLOWING CANCER SYMPTOMS: PAIN, SHORTNESS OF BREATH, LOSS OF WEIGHT AND EXTREME TIREDNESS, AND EMOTIONAL PROBLEMS SUCH AS ANXIETY OR DEPRESSION;

2) YOU ARE ABLE TO TAKE CARE OF YOURSELF WITHOUT HELP FROM OTHERS; YOU CAN WORK A NORMAL 40-HOUR WEEK; YOUR CONDITION DOES NOT INTERFERE WITH YOUR FAMILY LIFE OR WITH OTHER SOCIAL RELATIONS.

4. TO AVOID HAVING FINANCIAL DIFFICULTIES DUE TO YOUR CONDITION TWO YEARS FROM NOW.

5. TO AVOID THE ADVERSE EFFECTS DUE TO THE TREATMENT. THESE ARE: SHORTNESS OF BREATH, PROBLEMS SWALLOWING, INFECTIONS DUE TO A FALL IN YOUR DEFENCES, DIARRHOEA, VOMITING, LOSS OF HAIR, A PRICKLING OR BURNING SENSATION (ESPECIALLY IN YOUR HANDS OR FEET), FATIGUE, AND LOSS OF WEIGHT.

6. TO HAVE A GOOD EXPERIENCE OF THE HEALTH CARE RECEIVED IN THE NEXT TWO YEARS, YOU WILL HAVE A GOOD EXPERIENCE OF THE HEALTH CARE RECEIVED IF:

1) YOU HAVE TO ATTEND THE OUTPATIENT CLINIC AS FEW TIMES AS POSSIBLE;

2) YOU HAVE TO BE ADMITTED TO HOSPITAL AS FEW TIMES AS POSSIBLE

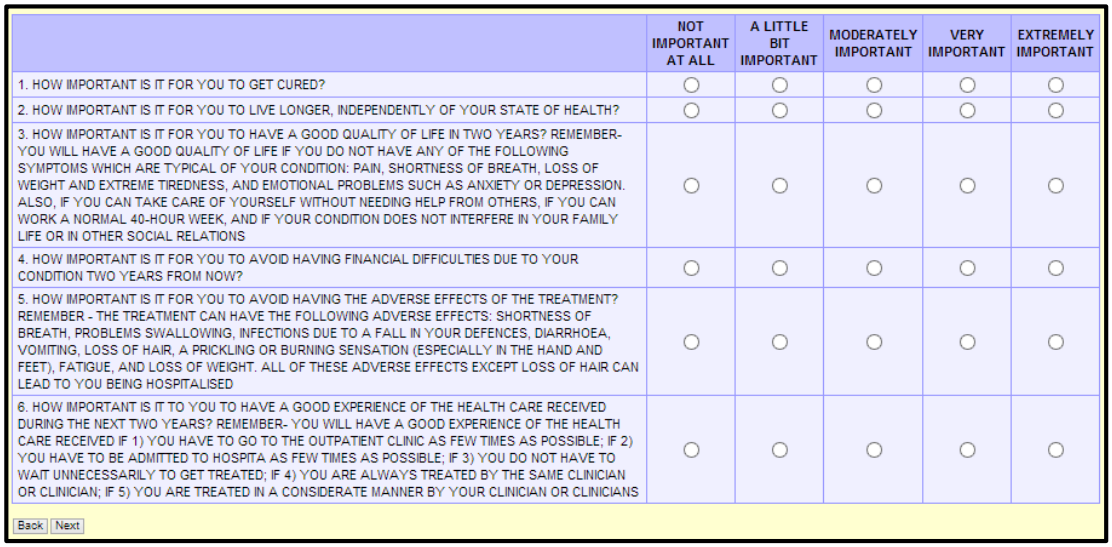
3) YOU DO NOT HAVE TO WAIT UNNECESSARILY TO GET TREATMENT

4) YOU ARE ALWAYS TREATED BY THE SAME CLINICIAN OR TEAM OF CLINICIANS

5) YOU ARE TREATED IN AN ATTENTIVE AND CONSIDERATE MANNER BY YOUR CLINICIAN OR CLINICIANS

Back Next

Figure A7.2. Elicia screen to assess the criteria weights



	NOT IMPORTANT AT ALL	A LITTLE BIT IMPORTANT	MODERATELY IMPORTANT	VERY IMPORTANT	EXTREMELY IMPORTANT
1. HOW IMPORTANT IS IT FOR YOU TO GET CURED?	○	○	○	○	○
2. HOW IMPORTANT IS IT FOR YOU TO LIVE LONGER, INDEPENDENTLY OF YOUR STATE OF HEALTH?	○	○	○	○	○
3. HOW IMPORTANT IS IT FOR YOU TO HAVE A GOOD QUALITY OF LIFE IN TWO YEARS? REMEMBER- YOU WILL HAVE A GOOD QUALITY OF LIFE IF YOU DO NOT HAVE ANY OF THE FOLLOWING SYMPTOMS WHICH ARE TYPICAL OF YOUR CONDITION: PAIN, SHORTNESS OF BREATH, LOSS OF WEIGHT AND EXTREME TIREDNESS, AND EMOTIONAL PROBLEMS SUCH AS ANXIETY OR DEPRESSION. ALSO, IF YOU CAN TAKE CARE OF YOURSELF WITHOUT NEEDING HELP FROM OTHERS, IF YOU CAN WORK A NORMAL 40-HOUR WEEK, AND IF YOUR CONDITION DOES NOT INTERFERE IN YOUR FAMILY LIFE OR IN OTHER SOCIAL RELATIONS	○	○	○	○	○
4. HOW IMPORTANT IS IT FOR YOU TO AVOID HAVING FINANCIAL DIFFICULTIES DUE TO YOUR CONDITION TWO YEARS FROM NOW?	○	○	○	○	○
5. HOW IMPORTANT IS IT FOR YOU TO AVOID HAVING THE ADVERSE EFFECTS OF THE TREATMENT? REMEMBER - THE TREATMENT CAN HAVE THE FOLLOWING ADVERSE EFFECTS: SHORTNESS OF BREATH, PROBLEMS SWALLOWING, INFECTIONS DUE TO A FALL IN YOUR DEFENCES, DIARRHOEA, VOMITING, LOSS OF HAIR, A PRICKLING OR BURNING SENSATION (ESPECIALLY IN THE HAND AND FEET), FATIGUE, AND LOSS OF WEIGHT. ALL OF THESE ADVERSE EFFECTS EXCEPT LOSS OF HAIR CAN LEAD TO YOU BEING HOSPITALISED	○	○	○	○	○
6. HOW IMPORTANT IS IT TO YOU TO HAVE A GOOD EXPERIENCE OF THE HEALTH CARE RECEIVED DURING THE NEXT TWO YEARS? REMEMBER- YOU WILL HAVE A GOOD EXPERIENCE OF THE HEALTH CARE RECEIVED IF 1) YOU HAVE TO GO TO THE OUTPATIENT CLINIC AS FEW TIMES AS POSSIBLE; IF 2) YOU HAVE TO BE ADMITTED TO HOSPITA AS FEW TIMES AS POSSIBLE; IF 3) YOU DO NOT HAVE TO WAIT UNNECESSARILY TO GET TREATED; IF 4) YOU ARE ALWAYS TREATED BY THE SAME CLINICIAN OR CLINICIAN; IF 5) YOU ARE TREATED IN A CONSIDERATE MANNER BY YOUR CLINICIAN OR CLINICIANS	○	○	○	○	○

Back Next

Figure A7.3. Weightings panel in the Annalisa topic screen

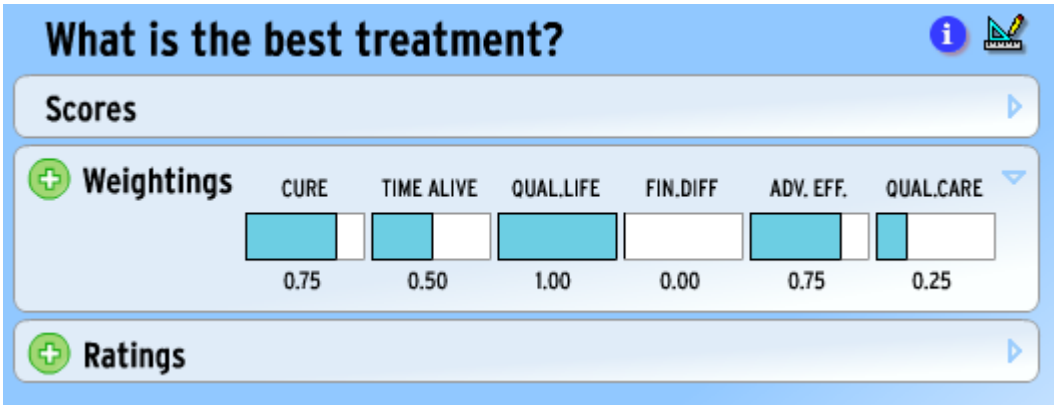


Figure A7.4. Information pop-up for criterion financial burden due to the disease in the medium term

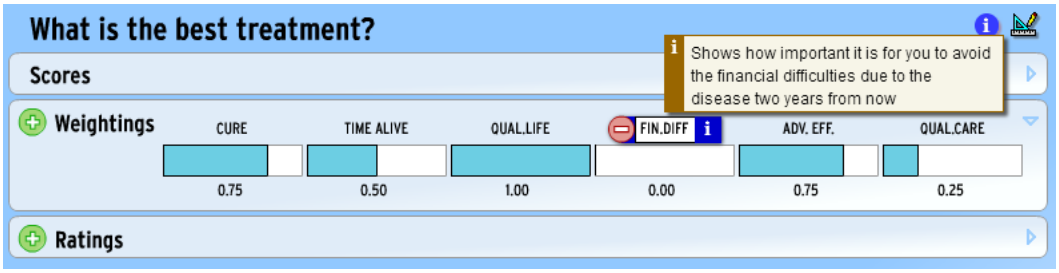
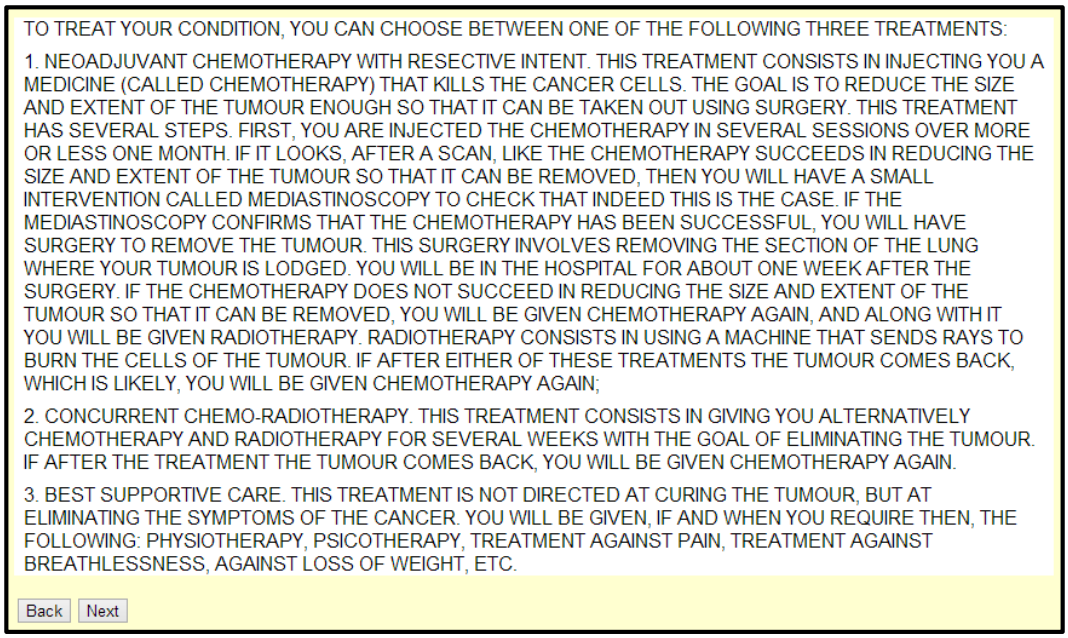


Figure A7.5. Communication of the options


TO TREAT YOUR CONDITION, YOU CAN CHOOSE BETWEEN ONE OF THE FOLLOWING THREE TREATMENTS:

1. NEOADJUVANT CHEMOTHERAPY WITH RESECTIVE INTENT. THIS TREATMENT CONSISTS IN INJECTING YOU A MEDICINE (CALLED CHEMOTHERAPY) THAT KILLS THE CANCER CELLS. THE GOAL IS TO REDUCE THE SIZE AND EXTENT OF THE TUMOUR ENOUGH SO THAT IT CAN BE TAKEN OUT USING SURGERY. THIS TREATMENT HAS SEVERAL STEPS. FIRST, YOU ARE INJECTED THE CHEMOTHERAPY IN SEVERAL SESSIONS OVER MORE OR LESS ONE MONTH. IF IT LOOKS, AFTER A SCAN, LIKE THE CHEMOTHERAPY SUCCEEDS IN REDUCING THE SIZE AND EXTENT OF THE TUMOUR SO THAT IT CAN BE REMOVED, THEN YOU WILL HAVE A SMALL INTERVENTION CALLED MEDIASTINOSCOPY TO CHECK THAT INDEED THIS IS THE CASE. IF THE MEDIASTINOSCOPY CONFIRMS THAT THE CHEMOTHERAPY HAS BEEN SUCCESSFUL, YOU WILL HAVE SURGERY TO REMOVE THE TUMOUR. THIS SURGERY INVOLVES REMOVING THE SECTION OF THE LUNG WHERE YOUR TUMOUR IS LODGED. YOU WILL BE IN THE HOSPITAL FOR ABOUT ONE WEEK AFTER THE SURGERY. IF THE CHEMOTHERAPY DOES NOT SUCCEED IN REDUCING THE SIZE AND EXTENT OF THE TUMOUR SO THAT IT CAN BE REMOVED, YOU WILL BE GIVEN CHEMOTHERAPY AGAIN, AND ALONG WITH IT YOU WILL BE GIVEN RADIOTHERAPY. RADIOTHERAPY CONSISTS IN USING A MACHINE THAT SENDS RAYS TO BURN THE CELLS OF THE TUMOUR. IF AFTER EITHER OF THESE TREATMENTS THE TUMOUR COMES BACK, WHICH IS LIKELY, YOU WILL BE GIVEN CHEMOTHERAPY AGAIN;

2. CONCURRENT CHEMO-RADIOTHERAPY. THIS TREATMENT CONSISTS IN GIVING YOU ALTERNATIVELY CHEMOTHERAPY AND RADIOTHERAPY FOR SEVERAL WEEKS WITH THE GOAL OF ELIMINATING THE TUMOUR. IF AFTER THE TREATMENT THE TUMOUR COMES BACK, YOU WILL BE GIVEN CHEMOTHERAPY AGAIN.

3. BEST SUPPORTIVE CARE. THIS TREATMENT IS NOT DIRECTED AT CURING THE TUMOUR, BUT AT ELIMINATING THE SYMPTOMS OF THE CANCER. YOU WILL BE GIVEN, IF AND WHEN YOU REQUIRE THEN, THE FOLLOWING: PHYSIOTHERAPY, PSICOTHERAPY, TREATMENT AGAINST PAIN, TREATMENT AGAINST BREATHLESSNESS, AGAINST LOSS OF WEIGHT, ETC.

Back Next

Figure A7.6. Communication of results

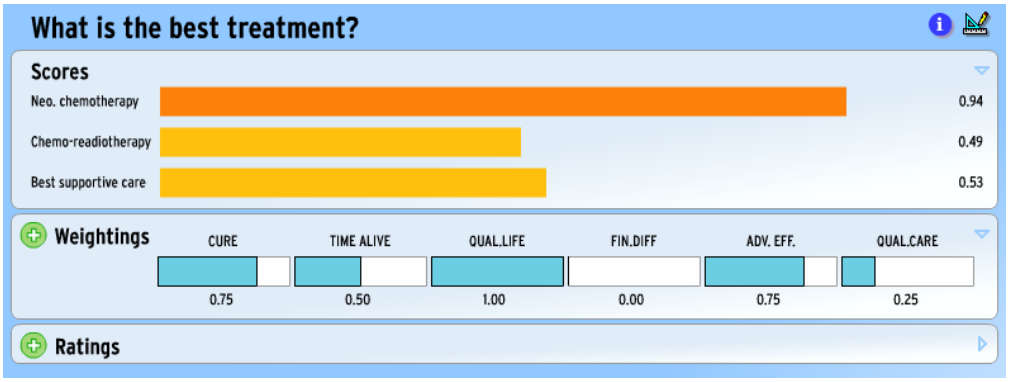


Figure A7.7. Communication of the evidence

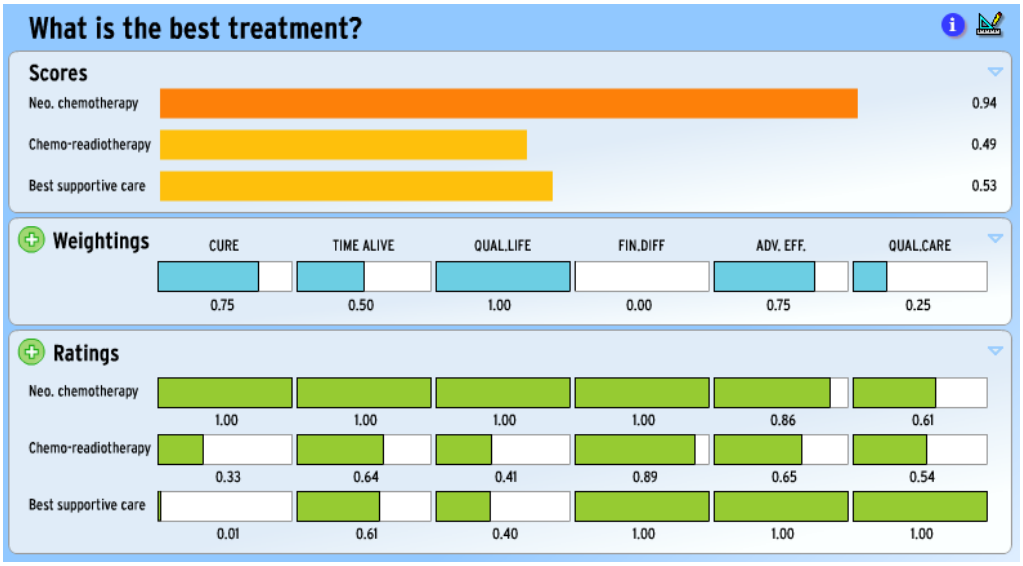
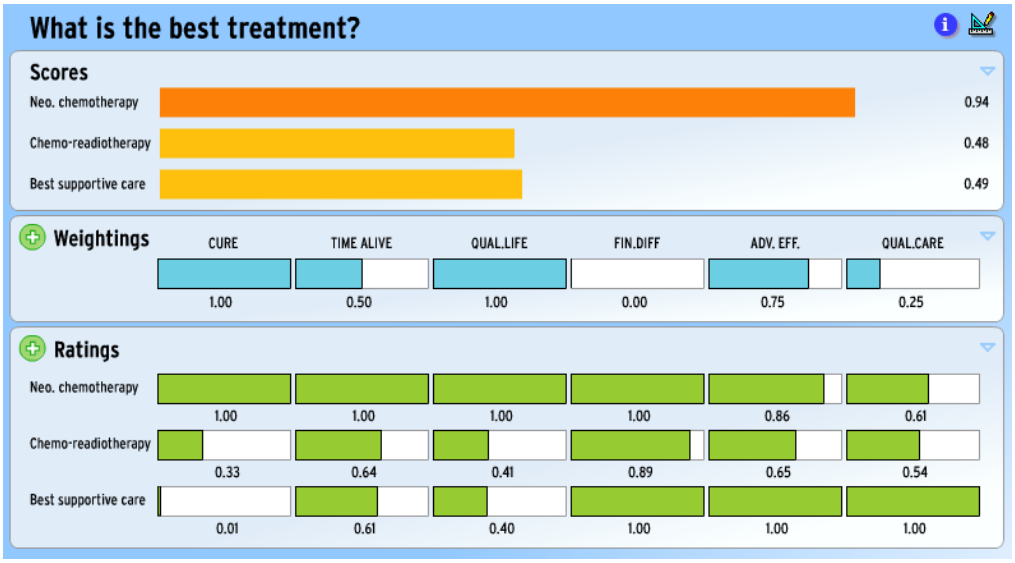


Figure A7.8 Sensitivity analysis of the criteria weights



Appendix 8: Spanish version of the top-level criteria and of the treatment descriptions for the *Expert Choice*-based PDA

Table A8.1. Renaming in Spanish the top-level criteria for communication to the patient

Top-level criterion	**Top-level criterion as it was renamed for communication to the patient**
Curarme	Maximizar las opciones de curarme
Esperanza de vida	Vivir el mayor tiempo que sea posible, independientemente de que mi estado de salud sea bueno o malo
Calidad de vida dentro de dos años	Tener la major calidad de vida possible dentro de dos años. Esto ocurrirá si: 1) Yo no tengo ninguno de los siguientes sintomas típicos del cáncer: dolor, sensación de ahogo, astenia, problemas emocionales como depresión o irritabilidad 2) Yo soy capaz de cuidar de mi mismo sin necesitar ayuda de otros, puedo trabajar una semana laboral normal, y mi enfermedad no interfiere de manera notable sobre mi vida familiar y social
Problemas económicos dentro de dos años	Tener los menores problemas económicos posibles por causa de la enfermedad dentro de dos años
Efectos adversos derivados del tratamiento	Padecer los menores efectos adversos derivados del tratamiento que sea posible. Entre los efectos adversos más communes están los siguientes: sensación de ahogo causada por la inflamación de un pulmón, problemas al tragar causados por una inflamación del esófago, infecciones causadas por una caída de las defensas, diarrea, vómitos, caída del pelo, picores y sensación de temblor en las extremidades, cansancio, anorexia. Todos estos efectos adversos con la excepción de la caída del pelo pueden llegar a requerir ingreso hospitalario
Calidad asistencial	Tener la mejor calidad asistencial posible durante los primeros dos años del tratamiento. Esto ocurrirá si: 1) yo no tengo que realizar visitas al hospital para seguir un tratamiento o realizar un chequeo, 2) yo no tengo que ser ingresado en el hospital, 3) yo no tengo que estar en lista de espera para recibir un tratamiento, 4) yo soy tratado siempre por el mismo médico o por el mismo equipo médico, 5) yo soy tratado en todo momento de forma personalizada y considerada

Table A8.2. Textual content in Spanish of the three treatment descriptions

Treatment option	**Description**
Neoadjuvant chemotherapy with resective intent (Option A_1)	1. Quimioterapia neoadyuvante con intención resectiva. Este tratamiento consiste en inyectarme una medicina (quimioterapia) que elimina las células del tumor para intentar reducir el tamaño de este y poder quitármelo con cirugía. El tratamiento tiene varias etapas: primero se me inyecta la quimioterapia en varias sesiones a lo largo de un mes. Si parece que la quimioterapia ha tenido éxito en reducir el tumor se me someterá a una pequeña intervención quirúrgica llamada mediastinoscopia para verificar que esto es así y que se me puede operar el tumor con éxito. Si se me puede operar el tumor, se me intervendrá (estaré ingresado aproximadamente una semana). Si la quimioterapia no ha tenido éxito en reducir el tumor se me volverá a dar quimioterapia a lo largo de unas semanas, esta vez combinada con radioterapia. La radioterapia consiste en el uso de una máquina que emite radiación para quemar las células del tumor. Si el tumor vuelve, cosa bastante probable, se me volverá a tratar con quimioterapia
Concurrent chemotherapy (Option A_2)	2. Quimio-radioterapia concomitante. Este tratamiento consiste en darme quimioterapia combinada con radioterapia a lo largo de unas semanas. Si el tumor vuelve, cosa bastante probable, se me volverá a tratar con quimioterapia
Best supportive care (Option A_3)	3. Tratamiento sintomático. Este tratamiento no va dirigido a eliminar el tumor, sino a tratar los síntomas de la enfermedad. Se me dará fisioterapia, psicoterapia, medicación para eliminar el dolor, para reducir los problemas para respirar, para reducir la náusea y los vómitos, para mejorar mi apetito, para eliminar el insomnio, etc

Appendix 9: Spanish version of the top-level criteria and of the treatment descriptions for the *ALEL*-based PDA

Table A9.1: Renaming in Spanish the top-level criteria for communication to the patient

Top-level criterion	**Top-level criterion as it was renamed for communication to the patient**
Curarme	Maximizar las opciones de curarme
Esperanza de vida	Vivir el mayor tiempo que sea posible, independientemente de que mi estado de salud sea bueno o malo
Calidad de vida dentro de dos años	Tener la major calidad de vida possible dentro de dos años. Esto ocurrirá si: 1) Yo no tengo ninguno de los siguientes sintomas típicos del cáncer: dolor, sensación de ahogo, astenia, problemas emocionales como depresión o irritabilidad 2) Yo soy capaz de cuidar de mi mismo sin necesitar ayuda de otros, puedo trabajar una semana laboral normal, y mi enfermedad no interfiere de manera notable sobre mi vida familiar y social
Problemas económicos dentro de dos años	Tener los menores problemas económicos posibles por causa de la enfermedad dentro de dos años
Efectos adversos derivados del tratamiento	Padecer los menores efectos adversos derivados del tratamiento que sea posible. Entre los efectos adversos más communes están los siguientes: sensación de ahogo causada por la inflamación de un pulmón, problemas al tragar causados por una inflamación del esófago, infecciones causadas por una caída de las defensas, diarrea, vómitos, caída del pelo, picores y sensación de temblor en las extremidades, cansancio, anorexia. Todos estos efectos adversos con la excepción de la caída del pelo pueden llegar a requerir ingreso hospitalario
Calidad asistencial	Tener la mejor calidad asistencial posible durante los primeros dos años del tratamiento. Esto ocurrirá si: 1) yo no tengo que realizar visitas al hospital para seguir un tratamiento o realizar un chequeo, 2) yo no tengo que ser ingresado en el hospital, 3) yo no tengo que estar en lista de espera para recibir un tratamiento, 4) yo soy tratado siempre por el mismo médico o por el mismo equipo médico, 5) yo soy tratado en todo momento de forma personalizada y considerada

Table A9.2: Textual content in Spanish of the three treatment descriptions

Treatment option	**Description**
Neoadjuvant chemotherapy with resective intent (Option A_1)	1. Quimioterapia neoadyuvante con intención resectiva. Este tratamiento consiste en inyectarme una medicina (quimioterapia) que elimina las células del tumor para intentar reducir el tamaño de este y poder quitármelo con cirugía. El tratamiento tiene varias etapas: primero se me inyecta la quimioterapia en varias sesiones a lo largo de un mes. Si parece que la quimioterapia ha tenido éxito en reducir el tumor se me someterá a una pequeña intervención quirúrgica llamada mediastinoscopia para verificar que esto es así y que se me puede operar el tumor con éxito. Si se me puede operar el tumor, se me intervendrá (estaré ingresado aproximadamente una semana). Si la quimioterapia no ha tenido éxito en reducir el tumor se me volverá a dar quimioterapia a lo largo de unas semanas, esta vez combinada con radioterapia. La radioterapia consiste en el uso de una máquina que emite radiación para quemar las células del tumor. Si el tumor vuelve, cosa bastante probable, se me volverá a tratar con quimioterapia
Concurrent chemotherapy (Option A_2)	2. Quimio-radioterapia concomitante. Este tratamiento consiste en darme quimioterapia combinada con radioterapia a lo largo de unas semanas. Si el tumor vuelve, cosa bastante probable, se me volverá a tratar con quimioterapia
Best supportive care (Option A_3)	3. Tratamiento sintomático. Este tratamiento no va dirigido a eliminar el tumor, sino a tratar los síntomas de la enfermedad. Se me dará fisioterapia, psicoterapia, medicación para eliminar el dolor, para reducir los problemas para respirar, para reducir la náusea y los vómitos, para mejorar mi apetito, para eliminar el insomnio, etc

Appendix 10: STATA output.

Panel 1. Assessment of inter-rater reliability for scores (priorities/ratings) of options A1, A2 and A3 on the six top-level criteria: 1) cure for cancer, 2) life expectancy, 3) quality of life in the medium term, 4) financial difficulties in the medium term, 5) adverse effects of treatment, 6) quality of the health care experience from start of treatment until the medium term

```
. bysort idHosp idTrat: icc punt20C idPunt idClin6
```

```
-> idHosp = 1, idTrat = 1

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |   .6678252       .4411757    .8370258
               Average |   .8577806       .7031244    .9390535
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 7.02              Prob > F = 0.000

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.
```

```
-> idHosp = 1, idTrat = 2

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |     .62123       .3812422     .809932
               Average |   .8310918       .6489286    .9274512
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 6.03              Prob > F = 0.000

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.

-------------------------------------------------------------------------------------------
-------------------------------------
-> idHosp = 1, idTrat = 3

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |   .6257861       .3835051    .8134242
               Average |   .8337988       .6511085    .9289735
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 5.91              Prob > F = 0.000

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.

-------------------------------------------------------------------------------------------
-------------------------------------
```

```
-> idHosp = 2, idTrat = 1

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |   .6768686       .4522445    .8423379
               Average |   .8627156       .7123872    .9412733
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 7.18              Prob > F = 0.000

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.

------------------------------------------------------------------------------------------------------------------------
-> idHosp = 2, idTrat = 2

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |   .8890574       .7848056    .9504473
               Average |   .9600655       .9162541    .9829181
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 26.32             Prob > F = 0.000

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.

------------------------------------------------------------------------------------------------------------------------
-> idHosp = 2, idTrat = 3

Intraclass correlations
Two-way random-effects model
Absolute agreement

Random effects: idPunt           Number of targets =        20
Random effects: idClin6          Number of raters  =         3

------------------------------------------------------------------
               punt20C |        ICC       [95% Conf. Interval]
-----------------------+------------------------------------------
            Individual |   .3446093       .0805348    .6222834
               Average |   .6120155       .2080878    .8317196
------------------------------------------------------------------
F test that
  ICC=0.00: F(19.0, 38.0) = 2.67              Prob > F = 0.005

Note: ICCs estimate correlations between individual measurements
      and between average measurements made on the same target.

.
```

Panel 2. Comparing the scores (priorities/ratings) of the options on the six top-level criteria across clinicians, options and criteria within each hospital.

1. Hospital H1.

```
regress punt6C i.IDTrat i.IDClin3 i.IDcrite if IDhosp==1 & IDTrat<3

      Source |       SS           df       MS      Number of obs   =
36
-------------+----------------------------------   F(8, 27)        =
11.64
       Model |  1.30381113         8  .162976391   Prob > F        =
0.0000
    Residual |  .378086126        27   .01400319   R-squared       =
0.7752
-------------+----------------------------------   Adj R-squared   =
0.7086
       Total |  1.68189726        35  .048054207   Root MSE        =
.11834

------------------------------------------------------------------------------
      punt6C |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    2.IDTrat |  -.2583333    .039445    -6.55   0.000    -.3392678   -
.1773988
             |
     IDClin3 |
          2  |   .0058333   .0483101     0.12   0.905    -.0932908
.1049575
          3  |      .0225   .0483101     0.47   0.645    -.0766241
.1216241
             |
     IDcrite |
          2  |       -.04   .0683208    -0.59   0.563    -.1801827
.1001827
          3  |  -.0016667   .0683208    -0.02   0.981    -.1418493
.138516
          4  |  -.0016667   .0683208    -0.02   0.981    -.1418493
.138516
          5  |      -.375   .0683208    -5.49   0.000    -.5151827   -
.2348173
          6  |       -.19   .0683208    -2.78   0.010    -.3301827   -
.0498173
             |
       _cons |   .6013889   .0591675    10.16   0.000     .4799871
.7227906
```

Wald test (Partial F)

```
. test (2.IDClin3 3.IDClin3)

 ( 1)  2.IDClin3 = 0
 ( 2)  3.IDClin3 = 0

       F(  2,    27) =    0.12
            Prob > F =    0.8902

. test (2.IDcrite 3.IDcrite 4.IDcrite 5.IDcrite 6.IDcrite )

 ( 1)  2.IDcrite = 0
 ( 2)  3.IDcrite = 0
 ( 3)  4.IDcrite = 0
 ( 4)  5.IDcrite = 0
 ( 5)  6.IDcrite = 0

       F(  5,    27) =   10.00
            Prob > F =    0.0000
```

2. Hospital H2.

```
. regress punt6C i.IDTrat i.IDClin3 i.IDcrite if IDhosp==2

      Source |       SS           df       MS      Number of obs   =
54
-------------+----------------------------------   F(9, 44)        =
4.13
       Model |  2.14989443         9  .238877159   Prob > F        =
0.0007
    Residual |  2.54318886        44  .057799747   R-squared       =
0.4581
-------------+----------------------------------   Adj R-squared   =
0.3473
       Total |  4.69308329        53  .088548741   Root MSE        =
.24042

------------------------------------------------------------------------------
      punt6C |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      IDTrat |
          2  |  -.2955556   .0801386    -3.69   0.001    -.4570643   -
.1340468
          3  |  -.3827778   .0801386    -4.78   0.000    -.5442865   -
.2212691
             |
     IDClin3 |
          2  |  -.0011111   .0801386    -0.01   0.989    -.1626198
.1603976
          3  |  -.0388889   .0801386    -0.49   0.630    -.2003976
.1226198
             |
     IDcrite |
          2  |        .16   .1133331     1.41   0.165    -.0684078
.3884078
          3  |   .1333333   .1133331     1.18   0.246    -.0950745
.3617412
          4  |        .28   .1133331     2.47   0.017     .0515922
.5084078
          5  |   .3122222   .1133331     2.75   0.009     .0838144
.54063
          6  |   .3044444   .1133331     2.69   0.010     .0760366
.5328523
             |
       _cons |   .7472222   .1034585     7.22   0.000     .5387154
.9557291
```

Wald test (partial F)

```
. test (2.IDTrat 3.IDTrat)

 ( 1)  2.IDTrat = 0
 ( 2)  3.IDTrat = 0

       F(  2,    44) =   12.53
            Prob > F =    0.0000

. test (2.IDClin3 3.IDClin3)

 ( 1)  2.IDClin3 = 0
 ( 2)  3.IDClin3 = 0

       F(  2,    44) =    0.15
            Prob > F =    0.8589

. test (2.IDcrite 3.IDcrite 4.IDcrite 5.IDcrite 6.IDcrite )

 ( 1)  2.IDcrite = 0
```

```
( 2)  3.IDcrite = 0
( 3)  4.IDcrite = 0
( 4)  5.IDcrite = 0
( 5)  6.IDcrite = 0

       F(  5,    44) =    2.36
            Prob > F =    0.0551
```

References:

1. Entwistle V. *Patient involvement in decision-making: The importance of a broad conceptualization*. In: Edwards A, Elwyn G, editors. Shared decision making in health care: Achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.

2. Edwards A, Elwyn G. *Shared decision-making in health care: Achieving evidence-based patient choice*. In: Edwards A, Elwyn G, editors. Shared decision-making in health care: Achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.

3. Elwyn G. *Implementing Shared Decision Making in the UK*. British Medical Journal. 2010;341:971-5.

4. Makoul G, Clayman ML. *An integrative model of shared decision making in medical encounters*. Patient Education and Counseling. 2006;60(3):301-12.

5. Stacey D, Legare F, et al. *Decision aids for people facing health treatment or screening decisions*. The Cochrane database of systematic reviews. 2014.

6. O'Connor A, Edwards A. *The role of decision aids in promoting evidence-based patient choice*. In: Edwards A, Elwyn G, editors. Shared decision making in health care: Achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.

7. O'Donnell S, Cranney A, et al. *Understanding and overcoming the barriers of implementing patient decision aids in clinical practice*. Journal of Evaluation in Clinical Practice. 2006;12(2):174-81.

8. Dolan JG, Boohaker E, et al. *Can Streamlined Multicriteria Decision Analysis Be Used to Implement Shared Decision Making for Colorectal Cancer Screening?* Medical Decision Making. 2013;34(6):746-55.

9. Belton V, Stewart TJ. *Multiple Criteria Decision Analysis: An integrated approach*. Dordrecht: Kluwer Academic Publishers; 2002.

10. Dodgson J, Spackman M, et al. *Multi-Criteria Analysis: a Manual*. London: Department of the Environment, Transport and the Regions; 2000 (republished 2009 by the Department for Communities and Local Government).

11. Keeney RL, Raiffa H. *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge: Cambridge University Press; 1993.

12. Saaty T. *The fundamentals of decision making and priority theory with the Analytic Hierarchy Process*. Pittsburgh: RWS; 2000.

13. Figueira J, Mousseau V, et al. *ELECTRE methods*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the Art Surveys. New York: Springer; 2005.

14. Brans J-P, Mareschal B. *PROMETHEE methods*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

15. *HIVIEW 3*. [May 5, 2015]; Available from: http://www.catalyzeconsulting.com/index.php/software/hiview3/.

16. *Expert Choice*. [May 6, 2015]; Available from: http://expertchoice.com/.

17. Adunlin G, Diaby V, et al. *Multicriteria decision analysis in oncology*. Health Expectations. Epub ahead of print 2014:http://dx.doi.org/10.1111/hex.12178.

18. Dolan JG. *Are patients capable of using the analytic hierarchy process and willing to use it to help make clinical decisions?* Medical Decision Making. 1995;15(1):76-80.

19. Dowie J, Kaltoft M, et al. *Towards generic online multicriteria decision support in patient-centred health care*. Health Expectations. Epub ahead of print 2013:http://dx.doi.org/10.1111/hex.12111.

20. *Annalisa in Elicia*. [May 13, 2015]; Available from: http://www.maldaba.co.uk/products/elicia/.

21. Kaltoft M, Cunich M, et al. *Assessing decision quality in patient-centred care requires a preference-sensitive measure*. Journal of Health Services Research and Policy. 2014;19(2):110-7.

22. van de Ven A, Johnson P. *Knowledge for theory and practice*. Academy of Management Review. 2006;31:802-21.

23. Holmes-Rovner M, Nelson WL, et al. *Are patient decision aids the best way to improve clinical decision making? Report of the IPDAS Symposium*. Medical Decision Making. 2007;27(5):599-608.

24. Charles C, Gafni A, et al. *Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango)*. Social Science and Medicine. 1997;44(5):681-92.

25. Garg AX, Adhikari NK, et al. *Effect of computerised clinical decision support systems on practitioner performance and patient outcomes*. Journal of the American Medical Association. 2005;293(10):1223-38.

26. Durand MA, Stiel M, et al. *Where is the theory? Evaluating the theoretical frameworks described in decision support technologies*. Patient Education and Counseling. 2008;71(1):125-35.

27. Pauker SP, Pauker SG. *The amniocentesis decision: an explicit guide for parents*. Birth Defects. 1979;15(5C):289-324.

28. Kasper JF, Mulley AG, et al. *Developing shared decision-making programs to improve the quality of health care*. Quality Review Bulletin. 1992;18(6):183-90.

29. Barry MJ, Fowler FJ, et al. *Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostate hyperplasia*. Medical Care. 1995;33(8):771-82.

30. Elwyn G, O'Connor A, et al. *Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process*. British Medical Journal. 2006;333(7565):417.

31. O'Connor A, Wennberg JE, et al. *Towards the "tipping point": Decision aids and informed patient choice*. Health Affairs. 2007;26(3):716-25.

32. Ottawa Hospital Research Institute. *A to Z Inventory of Patient Decision Aids* [January 20, 2015]; Available from: http://decisionaid.ohri.ca/azinvent.php.

33. O'Connor AM, Llewellyn-Thomas HA, et al. *Modifying unwarranted variations in health care: Shared decision making using patient decision aids*. Health Affairs (Millwood). 2004;Suppl Variation:VAR63-72.

34. Frosch DL, Kaplan RM. *Shared decision making in clinical medicine: Past research and future directions*. American Journal of Preventive Medicine. 1999;17:285-94.

35. Labrecque M, Paunescu C, et al. *Evaluation of the effect of a patient decision aid about vasectomy on the decision-making process a randomized trial*. Contraception. 2010;82(6):556-62.

36. Legare F, Labrecque M, et al. *Training family physicians in shared decision making for the use of antibiotics for acute respiratory infections: A pilot clustered randomized controlled trial*. Health Expectations. 2011;14:96-110.

37. Hunter AG, Cappelli M, et al. *A randomised trial comparing alternative approaches to prenatal diagnosis counseling in advanced maternal age patients*. Clinical Genetics. 2005;67(4):303-13.

38. Partin MR, Nelson D, et al. *Randomized trial examining the effect of two prostate cancer screening educational interventions on patient knowledge, preferences and behaviors*. Journal of General Internal Medicine. 2004;19(8):835-42.

39. Pignone M, Harris R, et al. *Videotape-based decision aid for colon cancer screening: A randomized, controlled trial*. Annals of Internal Medicine. 2000;133(10):761-9.

40. Patient. *Decision Aids*. [January 20, 2015]; Available from: http://www.patient.co.uk/decision-aids.

41. NHS Rightcare. *Patient Decision Aids*. [January 20, 2015]; Available from: http://sdm.rightcare.nhs.uk/pda/.

42. Healthwise. *Decision Aids*. [January 20, 2015]; Available from: http://www.healthwise.org/products/decisionaids.aspx.

43. Mayo Clinic Shared Decision Making National Resource Center. *Decision Aids*. [January 20, 2015]; Available from: http://shareddecisions.mayoclinic.org/.

44. Bekker HL, Hewison J, et al. *Understanding why decision aids work: Linking process with outcome*. Patient Education and Counseling. 2003;50(3):323-9.

45. Montgomery AA, Fahey T, et al. *A factorial randomised controlled trial of decision analysis and an information video plus leaflet for newly diagnosed hypertensive patients*. British Journal of General Practice. 2003;53(491):446-53.

46. Dolan JG, Frisina S. *Randomized controlled trial of a patient decision aid for colorectal cancer screening*. Medical Decision Making. 2002;22(2):125-39.

47. Lalonde L, O'Connor A, et al. *Evaluation of a decision aid and a personal risk profile in community pharmacy for patients considering options to improve cardiovascular health: The options pilot study*. International Journal of Pharmacy Practice. 2006;14(1):51-62.

48. Schapira MM, Vansuiswyk J. *The effect of an illustrated pamphlet decision-aid on the use of prostate cancer screening tests*. The Journal of Family Practice. 2000;49(5):418-24.

49. Auvinen A, Hakama A, et al. *A randomized trial of choice of treatment in prostate cancer: The effect of intervention on the treatment chosen*. BJU International. 2004;93(1):52-6.

50. Deschamps MA, Taylor JG, et al. *Impact of pharmacy consultation versus a decision aid on decision making regarding hormone replacement therapy*. International Journal of Pharmacy Practice. 2004;12(1):21-8.

51. Sheehan J, Sherman KA. *Computerised decision aids: A systematic review of their effectiveness in facilitating high-quality decision-making in various health-related contexts*. Patient Education and Counseling. 2012;88(1):69-86.

52. Kennedy A. *On what basis should the effectiveness of decision aids be judged?* Health Expectations. 2003;6(3):255-68.

53. Bennett C, Graham ID, et al. *Validation of a Preparation for Decision Making Scale*. Patient Education and Counseling. 2010;78(1):130-33.

54. Cleary PD. *Satisfaction May Not Suffice!: A Commentary on 'a Patient's Perspective'*. International Journal of Technology Assessment in Health Care. 1998;14(01):35-7.

55. Sepucha K, Fowler FJ, et al. *Policy support for patient-centered care: The need for measureable improvements in decision quality*. Health Affairs (Millwood). 2004;Suppl Variation:Var54-62.

56. Elwyn G, Hutchings H, et al. *The OPTION scale: Measuring the extent that clinicians involve patients in decision-making tasks*. Health Expectations. 2005;8(1):34-42.

57. Braddock CH, Fihn SD, et al. *How doctors and patients discuss routine clinical decisions: Informed decision-making in the outpatient setting*. Journal of General Internal Medicine. 1997;12(6):339-45.

58. Degner LF, Sloan JA, et al. *The Control Preferences Scale*. Canadian Journal of Nursing Research. 1997;29(3):21-43.

59. O'Connor A. *Validation of a decisional conflict scale*. Medical Decision Making. 1995;15(1):25-30.

60. O'Connor, A. *User Manual - Decisional Conflict Scale*. 2010 [February 20, 2015]; Available from: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf.

61. Sepucha KR, Borkhoff CM, et al. *Establishing the effectiveness of patient decision aids: Key constructs and measurement instruments*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S12.

62. O'Connor A, Pennie RA, et al. *Framing effects on expectations, decisions, and side effects experienced: The case of influenza immunization*. Journal of Clinical Epidemiology. 1996;49(11):1271-76.

63. Man-Son-Hing M, Laupacis A, et al. *A patient decision aid regarding antithrombotic therapy for stroke prevention in atrial fibrillation: A randomized controlled trial*. Journal of the American Medical Association. 1999;88(8):737-43.

64. Elwyn G, Miron-Shatz T. *Deliberation before determination: The definition and evaluation of good decision making*. Health Expectations. 2010;13(2):139-47.

65. Charles C, Gafni A, et al. *Treatment decision aids: Conceptual issues and future directions*. Health Expectations. 2005;8(2):114-25.

66. Volk R, Llewellyn-Thomas H, et al. *Ten years of the International Patient Decision Aid Standards Collaboration: Evolution of the core dimensions for assessing the quality of patient decision aids*. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):S1.

67. International Patient Decision Aids Collaboration. *IPDAS criteria for judging the quality of patient decision aids*. [January 20, 2015]; Available from: http://ipdas.ohri.ca/IPDAS_checklist.pdf.

68. Sepucha K, Mulley A. *A practical approach to measuring the quality of preference-sensitive decisions*. In: Edwards A, Elwyn G, editors. Shared decision-making in health care: Achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.

69. Health Decision Sciences Center, Massachusetts General Hospital. *Decision Quality Instruments*. [January 23, 2015]; Available from: http://www.massgeneral.org/decisionsciences/research/.

70. Entwistle VA, Sowden AJ, et al. *Evaluating interventions to promote patient involvement in decision-making: By what criteria should effectiveness be judged?* Journal of Health Services Research and Policy. 1998;3(2):100-7.

71. McCaffery K, Irwig L, et al. *Patient decision aids to support clinical decision making: Evaluating the decision or the outcomes of the decision*. Medical Decision Making. 2007;27(5):619-25.

72. Feldman-Stewart D, O'Brien MA, et al. *Providing information about options in patient decision aids*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S4.

73. Trevena LJ, Zikmund-Fisher BJ, et al. *Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S7.

74. Fagerlin A, Pignone M, et al. *Clarifying values: An updated review*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S8.

75. Llewellyn-Thomas H. *Values clarification*. In: Edwards A, Elwyn G, editors. Shared decision-making in health care: Achieving evidence-based patient choice. Oxford: Oxford University Press; 2009.

76. Gafni A. *The standard gamble: What is being measured and how it is interpreted*. Health Services Research. 1994;29(2):207-24.

77. Abhyankar P, Bekker HL, et al. *Why values elicitation techniques enable people to make informed decisions about cancer trial participation*. Health Expectations. 2011;14 Suppl 1:20-32.

78. O'Connor A, Tugwell P, et al. *Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy*. Medical Decision Making. 1998;18(3):295-303.

79. Attema AE, Edelaar-Peters Y, et al. *Time trade-off: One methodology, different methods*. European Journal of Health Economics. 2013;14(Suppl 1):53-64.

80. Hawley ST. *Conjoint analysis: A 'new' way to evaluate patients' preferences*. Patient. 2008;1(4):255-7.

81. Stacey D, Kryworuchko J, et al. *Coaching and guidance with patient decision aids: A review of theoretical and empirical evidence*. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):S11.

82. Abhyankar P, Volk RJ, et al. *Balancing the presentation of information and options in patient decision aids: An updated review*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S6.

83. Coulter A, Stilwell D, et al. *A systematic development process for patient decision aids*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S2.

84. Montori VM, LeBlanc A, et al. *Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: A quality dimension of the International Patient Decision Aid Standards*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S5.

85. Barry M, Chan E, et al. *Disclosing conflicts of interest in patient decision aids*. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):S3.

86. Bekker H, Winterbottom A, et al. *Do personal stories make patient decision aids more effective? A critical review of theory and evidence*. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):S9.

87. Bekker HL. *The loss of reason in patient decision aid research: Do checklists damage the quality of informed choice interventions?* Patient Education and Counseling. 2010;78(3):357-64.

88. Elwyn G, O'Connor AM, et al. *Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASi)*. PLoS One. 2009;4(3):e4705.

89. McDonald H, Charles C, et al. *Assessing the conceptual clarity and evidence base of quality criteria/standards developed for evaluating decision aids*. Health Expectations. 2014;17(2):232-43.

90. Elwyn G, Stiel M, et al. *The design of patient decision support interventions: Addressing the theory-practice gap*. Journal of Evaluation in Clinical Practice. 2011;17(4):565-74.

91. Kahneman D, Tversky A. *Prospect Theory: An analysis of decisions under risk*. Econometrica. 1979;47(2):263-91.

92. Reyna V. *A theory of medical decision-making and health: Fuzzy Trace Theory*. Medical Decision Making. 2008;28(6):850-65.

93. Anderson CJ. *The psychology of doing nothing: Forms of decision avoidance result from reason and emotion*. Psychological Bulletin. 2003;129(1):139-67.

94. Joseph-Williams N, Newcombe RG, et al. *Toward minimum standards for certifying patient decision aids: A modified delphi consensus process*. Medical Decision Making. 2014;34(6):699-710.

95. Michie S, Dormandy E, et al. *The Multidimensional Measure of Informed Choice: A validation study*. Patient Education and Counseling. 2002;48(1):87-91.

96. Hoffman AS, Volk RJ, et al. *Delivering patient decision aids on the Internet: Definitions, theories, current evidence, and emerging research areas*. BMC Medical Informatics and Decision Making. 2013;13 Suppl 2:S13.

97. Jones LA, Weymiller AJ, et al. *Should clinicians deliver decision aids? Further exploration of the statin choice randomized trial results*. Medical Decision Making. 2009;29(4):468-74.

98. Visschers VH, Meertens RH, et al. *Probability information in risk communication: A review of the research literature*. Risk Analysis. 2009;29(2):267-87.

99. Gigerenzer G, Hoffrage U. *How to improve bayesian reasoning without instruction: Frequency formats*. Psychological Review.102(4):684-704.

100. Han PK, Klein WM, et al. *Representing randomness in the communication of indicidualised cancer risk estimates: Effects on cancer risk perceptions, worry, and subjective uncertainty about risk*. Patient Education and Counseling. 2012;86(1):106-13.

101. Han PK, Klein WM, et al. *Communication of uncertainty regarding individualized cancer risk estimates: Effects and influential factors*. Medical Decision Making. 2011;31(2):354-66.

102. Witteman HO, Scherer LD, et al. *Values clarification exercises: A systematic review*. Society for Medical Decision Making: Advanced Designing of Evidence-Based Patient Decision Aids; Phoenix (Arizona) 2012.

103. Davidson B, Joyce RN, et al. *Empowerment of men newly diagnosed with prostate cancer*. Cancer Nursing. 1997;20(3):187-96.

104. Kennedy AD, Sculpher MJ, et al. *Effects of decision aids for menorrhagia on treatment choices, health outcomes, and costs*. Journal of the American Medical Association. 2002;288(21):2701-09.

105. Elwyn G, Scholl I, et al. *"Many miles to go …": A systematic review of the implementation of patient decision support interventions into routine clinical practice*. BMC Medical Informatics and Decision Making. 2013;13.

106. Graham ID, Logan J, et al. *A qualitative study of physicians' perceptions of three decision aids*. Patient Education and Counseling. 2003;50(3):279-83.

107. Legare F, Witteman HO. *Shared decision making: Examining key elements and barriers to adoption into routine clinical practice*. Health Affairs (Millwood). 2013;32(2):276-84.

108. Legare F, Ratte S, et al. *Interventions for improving the adoption of shared decision making by healthcare professionals*. Cochrane Database of Systematic Reviews. 2010(5).

109. Legare F, Politi MC, et al. *Training health professionals in shared decision-making: An international environmental scan*. Patient Education and Counseling. 2012;88(2):159-69.

110. Brackett C, Kearing S, et al. *Strategies for distributing cancer screening decision aids in primary care*. Patient Education and Counseling.78(2):166-8.

111. Legare F, Ratte S, et al. *Barriers and facilitators to implementing shared decision-making in clinical practice: Update of a systematic review of health professionals' perceptions*. Patient Education and Counseling. 2008;73(3):526-35.

112. McCaffery K, Holmes-Rovner M, et al. *Addressing health literacy in patient decision aids*. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):S10.

113. Ishizaka A, Nemery P. *Multi-Criteria Decision Analysis: Methods and software*. Chichester: John Wiley and Sons; 2013.

114. Figueira J, Greco S, et al., editors. *Multiple Criteria Decision Analysis: State of the art surveys*. New York: Springer; 2005.

115. Koksalan M, Wallenius J, et al. *Multiple Criteria Decision Making: From early history to the 21st Century*. London: World Scientific Publishing; 2011.

116. Dolan JG. *Multi-criteria clinical decision support: A primer in the use of multiple-criteria decision-making methods to promote evidence-based, patient-centered health care*. Patient. 2010;3(4):229-48.

117. Kilgour DM, Chen Y, et al. *Multiple Criteria approaches and group decision and negotiation*. In: Ehrgott M, Figueira J, Greco S, editors. Trends in Multiple Criteria Decision Analysis. New York: Springer; 2010.

118. Keeney RL. *Value-focused thinking: A path to creative decisionmaking*. Cambridge: Harvard University Press; 1992.

119. Sanderson C, Gruen R. *Analytical models for decision making*. Maidenhead: Open University Press; 2006.

120. Tversky A. *Elimination by aspects: A theory of choice*. Psychological Review. 1972;79(4).

121. Edwards W, Barron FH. *SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement*. Organizational Behavior and Human Decision Processes. 1994;60(3):306-25.

122. Saaty TL, Vargas LG. *Models, methods, concepts and applications of the Analytic Hierarchy Process*. New York: Springer; 2012.

123. Saaty TL. *The Analytic Network Process*. Pittsburgh: RWS Publications; 2001.

124. Bana e Costa CA. *On the mathematical foundations of MACBETH*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

125. Siskos Y. *UTA methods*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

126. Roy B. *Paradigms and challenges*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

127. Martel J-M, Matarao B. *Other outranking approaches*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

128. Jones D, Tamiz M. *Practical goal programming*. New York: Springer; 2010.

129. Behzadian M, Khanmohammadi Otaghsara S, et al. *A state-of the-art survey of TOPSIS applications*. Expert Systems with Applications. 2012;39(17):13051-69.

130. Korhonen P. *A visual reference direction approach to solving discrete multiple criteria problems*. European Journal of Operational Research. 1988;34(2):152-9.

131. Wierzbicky AP. *The need for and possible methods of objective ranking*. In: Ehrgott M, Figueira J, Greco S, editors. Trends in Multiple Criteria Decision Analysis. New Yotk: Springer; 2010.

132. Kahraman C, editor. *Fuzzy Multi-Criteria Decision Making: Theory and applications with recent developments*. New York: Springer; 2008.

133. Bellman RE, Zadeh LA. *Decision-Making in a Fuzzy Environment*. Management Science. 1970;17(4):B141-64.

134. Baas SM, Kwakernaak H. *Rating and ranking of multiple-aspect alternatives using fuzzy sets*. Automatica. 1977;13(1):47-58.

135. Ribeiro R. *Fuzzy multiple attribute decision making: A review and new preference elicitation techniques*. Fuzzy Sets and Systems. 1996;78(2):155-81.

136. Demirel T, Demirel NC, et al. *Fuzzy Analytic Hierarchy Process and its application*. In: Kahraman C, editor. Fuzzy Multi-Criteria Decision Making: Theory and application with recent developments. New York: Springer; 2008.

137. Bufardi A, Gheorghe R, et al. *Fuzzy outranking methods: Recent developments*. In: Kahraman C, editor. Fuzzy Multi-Criteria Decision Making: Theory and applications with recent developments. New York: Springer; 2008.

138. Jiménez A, Mateos A, et al. *Dominance intensity measure within fuzzy weight oriented MAUT: An application*. Omega. 2013;41(2):397-405.

139. Lahdelma R, Salminen P. *Stochastic Multicriteria Acceptability Analysis*. In: Ehrgott M, Figueira J, Greco S, editors. Trends in Multiple Criteria Decision Analysis. New York: Springer; 2010.

140. Lahdelma R, Salminen P. *SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making*. Operations Research. 2001;49(3):444-54.

141. Greco S, Matarazzo B. *Decision rule approach*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

142. Moshkovich H, Mechitov A, et al. *Verbal decision analysis*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

143. Larichev OJ, Moshkovich H. *ZAPROS-LM: A method and system for ordering multiattribute alternatives*. European Journal of Operational Research. 1995;82(3):503-21.

144. Adunlin G, Diaby V, et al. *Application of multicriteria decision analysis in health care: A systematic review and bibliometric analysis*. Health Expectations. 2014:Epub ahead of print 2014: :http://dx.doi.org/10.1111/hex.12287.

145. Dyer J. *Multi-Attribute Utility Theory*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

146. Riabacke M, Danielson M, et al. *State-of-the-Art Prescriptive Criteria Weight Elicitation*. Advances in Decision Sciences. 2012;2012:1-24.

147. Dyer J, Sarin R. *Measurable multiattribute value functions*. Operations Research. 1979;27(4):810-22.

148. *V.I.S.A. Decisions*. [May 7, 2015]; Available from: http://www.visadecisions.com/.

149. *Logical Decisions*. [May 7, 2015]; Available from: http://www.logicaldecisions.com/.

150. Drummond M, Sculpher M, et al. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2005.

151. von Winterfeldt D, Edwards W. *Decision analysis and behavioral research*. Cambridge: Cambridge University Press; 1986.

152. *GMAA*. [May 12, 2015]; Available from: http://www.dia.fi.upm.es/~ajimenez/GMAA.

153. Bana e Costa CJ, de Corte J, et al. *MACBETH. LSE OR working paper*. Available from: http://eprints.lse.ac.uk/22761/1/MACBETH_LSE_working_paper_0356_30set.pdf.

154. *M-MACBETH*. [May 5, 2015]; Available from: http://www.m-macbeth.com/en/m-home.html.

155. French S, Rios Insua D. *Statistical decision theory*. Oxford: Oxford University Press; 2000.

156. Lenert LA, Treadwell JR. *Effects on preferences of violations of procedural invariance*. Medical Decision Making. 1999;19(4):473-81.

157. Johnson EM, Huber GP. *The technology of utility assessment*. Systems, Man and Cybernetics, IEEE Transactions on. 1977;7(5):311-25.

158. Belton V. *A comparison of the analytic hierarchy process and a simple multi-attribute value function*. European Journal of Operational Research. 1986;26:7-21.

159. Monat JP. *The benefits of global scaling in multi-criteria decision analysis*. Judgment and Decision Making. 2009;4(6):492-508.

160. van Til JA, Dolan JG, et al. *The use of multi-criteria decision analysis weight elicitation techniques in patients with mild cognitive impairment: a pilot study*. Patient-Patient Centered Outcomes Research. 2008;1(2):127-35.

161. Chang KY, Chan KH, et al. *Decision analysis for epidural labor analgesia with Multiattribute Utility (MAU) model*. The Clinical Journal of Pain. 2008;24(3):265-72.

162. Pinheiro PR, Castro A, et al., editors. *A multicriteria model applied in the diagnosis of Alzheimer's disease: a Bayesian network*. Computational Science and Engineering, 2008 CSE'08 11th IEEE International Conference on; 2008: IEEE.

163. Shaw RW, Brickley MR, et al. *Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment*. British Journal of Obstetrics and Gynaecology. 1998;105(11):1155-9.

164. de Bock GH, Reijneveld SA, et al. *Multiattribute utility scores for predicting family physicians' decisions regarding sinusitis*. Medical Decision Making. 1999;19(1):58-65.

165. Bettinger TL, Shuler G, et al. *Schizophrenia: Multi-attribute utility theory approach to selection of atypical antipsychotics*. Annals of Pharmacotherapy. 2007;41(2):201-7.

166. Suehs BT, Bettinger TL. *A multiattribute decision model for bipolar disorder: Identification of preferred mood-stabilizing medications*. Am J Manag Care. 2009;15(7):e42-52.

167. Schwartz MD, Valdimarsdottir HB, et al. *Randomized trial of a decision aid for BRCA1/BRCA2 mutation carriers: Impact on measures of decision making and satisfaction*. Health Psychology. 2009;28(1):11-9.

168. Hooker GW, Leventhal KG, et al. *Longitudinal changes in patient distress following interactive decision aid use among BRCA1/2 carriers: A randomized trial*. Medical Decision Making. 2011;31(3):412-21.

169. Hwang C-L, Yoon K. *Multiple attribute decision making. Methods and applications: A state-of-the-art survey*. New York: Springer-Verlag; 1981.

170. Eckenrode RT. *Weighting multiple criteria*. Management Science. 1965;12(3):180-92.

171. *Elicia*. [May 12, 2015]; Available from: http://www.elicia.org.uk/.

172. Wang Y-M, Luo Y. *On rank reversal in decision analysis*. Mathematical and Computer Modelling. 2009;49(5–6):1221-9.

173. Van Wijk BL, Klungel OH, et al. *A comparison of two multiple-characteristic decision-making models for the comparison of antihypertensive drug classes: Simple Additive Weighting (SAW) and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS)*. American Journal of Cardiovascular Drugs. 2006;6(4):251-8.

174. Azar FS. *Multiattribute decision-making: Use of three scoring methods to compare the performance of imaging techniques for breast cancer detection*. Technical Reports (CIS) Paper 119. 2000.

175. Masya LM, Young JM, et al. *Preferences for outcomes of treatment for rectal cancer: Patient and clinical utilities and their application in an interactive computer-based decision aid*. Diseases of the Colon and Rectum. 2009;52(12):1994-2003.

176. Cunich M, Salkeld G, et al. *Integrating evidence and individual preferences using a web-based multi-criteria decision analytic tool*. The Patient: Patient-Centered Outcomes Research. 2011;4(3):153-62.

177. Kaltoft MK, Dowie J, et al. *Addressing the disconnect between public health science and personalised health care: The potential role of cluster analysis in combination with multi-criteria decision analysis*. The Lancet. 2013;382:S52.

178. Dolan J, Isselhardt B, et al. *The Analytic Hierarchy Process in medical decision making: A tutorial*. Medical Decision Making. 1989;9:40-50.

179. Dolan J. *Involving patients in decisions regarding preventive health interventions using the Analytic Hierarchy Process*. Health Expectations. 2000;3:37-45.

180. Harker PT, Vargas LG. *The theory of ratio scale estimation: Saaty's Analytic Hierarchy Process*. Management Science. 1987;33(11):1383-403.

181. *MakeItRational*. [May 6, 2015]; Available from: http://makeitrational.com/.

182. Saaty TL. *Fundamentals of the Analytic Network Process — Dependence and feedback in decision-making with a single network*. Journal of Systems Science and Systems Engineering. 2004;13(2):129-57.

183. *Superdecisions*. [May 7, 2015]; Available from: http://www.superdecisions.com/.

184. Saaty TL. *Axiomatic foundation of the Analytic Hierarchy Process*. Management Science. 1986;32(7):841-55.

185. Dyer JS. *Remarks on the Analytic Hierarchy Process*. Management Science. 1990;36(3):249-58.

186. Saaty TL. *Reply to "Remarks on the Analytic Hierarchy Process"*. Management Science. 1990;36(3):269-73.

187. Salo AA, Hamalainen RP. *On the measurement of preferences in the Analytic Hierarchy Process*. Journal of Multi-Criteria Decision Analysis. 1997;6:309-19.

188. Belton V, Gear T. *On a short-coming of Saaty's method of analytic hierarchies*. Omega. 1983;11(3):228-30.

189. Johnson CR, Beine WB, et al. *Right-left asymmetry in an eigenvector ranking procedure*. Journal of Mathematical Psychology. 1979;19(1):61-4.

190. Schenkerman S. *Avoiding rank reversal in AHP decision-support models*. European Journal of Operational Research. 1994;74(3):407-19.

191. Saaty TL. *Decision making with the Analytic Hierarchy Process*. International Journal of Services Sciences. 2008;1(1):83-98.

192. Saaty TL. *The Analytic Hierarchy and Analytic Network Processes*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

193. Dolan JG. *Patient priorities in colorectal cancer screening decisions*. Health Expectations. 2005;8(4):334-44.

194. Liberatore MJ, Myers RE, et al. *Decision counseling for men considering prostate cancer screening*. Computers & OR. 2003;30(10):1421-34.

195. Carter KJ, Ritchey NP, et al. *Analysis of three decision-making methods: A breast cancer patient as a model*. Medical Decision Making. 1999;19(1):49-57.

196. Corrente S, Greco S, et al. *Multiple Criteria Hierarchy Process with ELECTRE and PROMETHEE*. Omega. 2013;41(5):820-46.

197. *ELECTRE Software*. [May 17, 2015]; Available from: http://www.lamsade.dauphine.fr/spip.php?article558.

198. *Visual PROMETHEE*. [May 10, 2015]; Available from: http://www.promethee-gaia.net/software.html.

199. *D-Sight*. [May 7,2015]; Available from: http://www.d-sight.com/?gclid=CK6E65mz5MUCFc_HtAodlBcA4g.

200. *Smart Picker Pro*. [May 13, 2015]; Available from: http://www.smart-picker.com/.

201. Le Gales C, Moatti JP. *Searching for consensus through multi-criteria decision analysis. Assessment of screening strategies for hemoglobinopathies in southeastern France*. International Journal of Technology Assessment in Health Care. 1990;6(3):430-49.

202. Brasil Filho A, Pinheiro P, et al. *Towards the Early Diagnosis of Alzheimer's Disease via a Multicriteria Classification Model*. In: Ehrgott M, Fonseca C, Gandibleux X, Hao J-K, Sevaux M, editors. Evolutionary Multi-Criterion Optimization: Springer Berlin Heidelberg; 2009. p. 393-406.

203. *Tryptich*. [May 15, 2015]; Available from: http://www.stat-design.com/Software/TOPSIS.html.

204. García-Cascales MS, Lamata MT. *On rank reversal and TOPSIS method*. Mathematical and Computer Modelling. 2012;56(5–6):123-32.

205. Ozernoy VM. *Choosing the best multiple criteria decision-making method*. INFOR. 1992;30(2):159-71.

206. Zanakis SH, Solomon A, et al. *Multi-attribute decision making: A simulation comparison of select methods*. European Journal of Operational Research. 1998;107(3):507-29.

207. Buede DM, Maxwell DT. *Rank disagreement: A comparison of multi-criteria methodologies*. Journal of Multi-Criteria Decision Analysis. 1995;4(1):1-21.

208. Salminen P, Hokkanen J, et al. *Comparing multicriteria methods in the context of environmental problems*. European Journal of Operational Research. 1998;104(3):485-96.

209. Mysiak J. *Consistency of the results of different MCA methods: a critical review*. Environment and Planning C-Government and Policy. 2006;24(2):257-77.

210. Hobbs BF. *What Can We Learn from Experiments in Multiobjective Decision Analysis?* Systems, Man and Cybernetics, IEEE Transactions on. 1986;16(3):384-94.

211. Guitouni A, Martel J-M. *Tentative guidelines to help choosing an appropriate MCDA method*. European Journal of Operational Research. 1998;109(2):501-21.

212. Phillips L. *A theory of requisite decision models*. Acta Psychologica. 1984;53:29-48.

213. Weistroffer HR, Smith CH, et al. *Multiple criteria decision support software*. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State of the art surveys. New York: Springer; 2005.

214. OR/MS Today. *Decision Analysis Software Survey, September 2014*. [May 5, 2015]; Available from: http://www.orms-today.org/surveys/das/das.html.

215. *MCDM software*. [May 5, 2015]; Available from: http://www.mcdmsociety.org/soft.html.

216. *Web-HIPRE*. [May 7, 2015]; Available from: http://hipre.aalto.fi/.

217. Mustajoki J, Hamalainen RP. *Web-Hipre: Global decision support by value tree and AHP Analysis* INFOR. 2000;38(3):208-20.

218. *PUrE2 Software*. [May 15, 2015]; Available from: http://www.pureintrawise.org/.

219. *MCDA-Res*. [May 15, 2015]; Available from: http://www.env.aegean.gr/energy/mcda/MCDA-toolkit.html.

220. Mustajoki J, Marttunen M. *Comparison of multi-criteria decision analytical software: Searching for ideas for developing a new EIA-specific multi-criteria software*: Finnish Environment Institute; 2013.

221. *Analytica*. [May 10, 2015]; Available from: http://www.lumina.com/why-analytica/.

222. *Criterium Decision Plus*. [May 5, 2015]; Available from: http://www.infoharvest.com/ihroot/infoharv/products.asp.

223. *OnBalance*. [May 8, 2015]; Available from: http://www.quartzstar.com/software.html.

224. *Smart Decisions*. [May 2, 2015]; Available from: http://www.cogentus.co.uk/fr/products/.

225. French S, Xu D-L. *Comparison study of multi-attribute decision analytic software*. Journal of Multi-Criteria Decision Analysis. 2005;13:65-80.

226. Baizildayeva U, Vlasov O, et al. *Multi-Criteria decision support systems. Comparative analysis*. Middle-East Journal of Scientific Research. 2013;16(12):1725-30.

227. Phillips LD, editor. *Decision conferencing*. CSCW: Some Fundamental Issues, IEE Colloquium on; 1991 15 Mar 1991.

228. Papamichail KN, French S. *Explaining and justifying the advice of a decision support system: A natural language generation approach*. Expert Systems with Applications. 2003;24(1):35-48.

229. Hodgkin J, Belton V, et al. *Supporting the intelligent MCDA user: A case study in multi-person multi-criteria decision support*. European Journal of Operational Research. 2005;160(1):172-89.

230. Jadhav AS, Sonar RM. *Evaluating and selecting software packages: A review*. Information and Software Technology. 2009;51:555-63.

231. Le Blanc LA, Tawfik Jelassi M. *DSS software selection: A multiple criteria decision methodology*. Information & Management. 1989;17(1):49-65.

232. Ossadnik W, Lange O. *AHP-based evaluation of AHP-Software*. European Journal of Operational Research. 1999;118(3):578-88.

233. Phillips-Wren GE, Hahn ED, et al. *A multiple-criteria framework for evaluation of decision support systems*. Omega. 2004;32(4):323-32.

234. Sobin LM, Gosporadovicz M, et al., editors. *TNM classification of malignant tumours: Seventh edition*. Oxford: Wiley-Blackwell; 2009.

235. Lopez-Encuentra A. *Carcinoma broncogenico: Pauta clinica*. Madrid: Hospital 12 de Octubre; 2010.

236. Dolan JG. *Medical decision making using the analytic hierarchy process: Choice of initial antimicrobial therapy for acute pyelonephritis*. Medical Decision Making. 1989;9(1):51-6.

237. *NCCN Clinical Practice Guidelines in Oncology: Non-Small Cell Lung Cancer*: National Comprehensive Cancer Network; 2010.

238. *Guias de practica clinica sobre cuidados paliativos*: Agencia de Evaluacion de Tecnologias Sanitarias del Pais Vasco ; 2008.

239. Aaronson N, Ahmedsai S, et al. *The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology*. Journal of the National Cancer Institute. 1993;85(5):365-76.

240. *The EORTC QLQ-C30 Scoring Manual*. Brussels: 2001 [January 2, 2015]; Available from: http://groups.eortc.be/qol/manuals.

241. Trotti A, Colevas AD, et al. *CTCAE v3.0: Development of a comprehensive grading system for the adverse effects of cancer treatment*. Seminars in Radiation Oncology.13(3):176-81.

242. *National Cancer Institute Therapy Evaluation Programme: Common Terminology Criteria for Adverse Effects Version 4 (CTCAE v4)*. [January 2, 2015]; Available from: http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm.

243. *Programa de Atencio Domiciliaria i Equips de Suport (PADES), CatSalut*. [May 12, 2015]; Available from: http://www.mutuam.com/sanitat/pades/.

244. Schrout PE, Fleiss JL. *Intraclass correlations: Uses in assessing rater reliability*. Psychological Bulletin. 1979;86(2):420-28.

245. *STATA 14*. [May 12, 2015]; Available from: http://www.stata.com/.

246. Landis JR, Kock GG. *The measurement of observer agreement for categorical data*. Biometrics. 1977;33:159-74.

247. Fleiss JL, Cohen J. *The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability*. Educational and Psychological Measurement. 1973;33:613-19.

248. *RETICS*. [May 12, 2015]; Available from: http://www.isciii.es/ISCIII/es/contenidos/fd-investigacion/fd-ejecucion/fd-centros-participados/centros-participados-redes-retics.shtml.

249. *CIBER*. [May 12, 2015]; Available from: http://www.isciii.es/ISCIII/es/contenidos/fd-investigacion/fd-ejecucion/fd-centros-participados/fd-consorcios2/cibers.shtml.

250. Kon AA. *The shared decision-making continuum*. Journal of the American Medical Association. 2010;304(8):903-4.

251. *Phillips, L. The mathematics of HiView and Equity*. 2004; Available from: http://www.catalyze.co.uk/wp-content/uploads/2012/07/Catalyze_White_Paper_Mathematics_of_Hiview_and_Equity.pdf.

252. Schroen A, Detterbeck FC, et al. *Beliefs among pulmonologists and thoracic surgeons in the therapeutic approach to non-small cell lung cancer*. Chest. 2000;118:129-37.

253. Camerer C, Johnson E. *The process performance paradox in expert judgment: How can experts know so much and predict so poorly?* In: Ericcsson K, Smith J, editors. Towards a General Theory of Expertise: Prospects and Limits. Cambridge: Cambridge University Press; 1991.

254. Lienard A, Merckaert I, et al. *Factors that influence cancer patient's anxiety following a medical consultation: Impact of a communication skills training programme for physicians*. Annals of Oncology. 2006;17(9):1450-58.

255. Whelan T, Sawka CA, et al. *Helping patients make informed choices: A randomized trial of a decision aid for adjuvant chemotherapy in lymph node-negative breast cancer*. Journal of the Lung Cancer Institute. 2003;95(8):581-87.

256. Bekker H, Legare F, et al. *Is anxiety a suitable measure of decision aid effectiveness? A systematic review*. Patient Education and Counseling. 2003;50:255-62.

257. Barry M, Cherkin D, et al. *A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia*. Disease Management and Clinical Outcomes. 1997;1(1):5-14.

258. Deyo RA, Cherki DC, et al. *Involving patients in clinical decisions: Impact of an interactive video program on use of back surgery*. Medical Care. 2000;38(9):959-69.

259. Murray E, Davis H, et al. *Randomised controlled trial of an interactive multimedia decision aid on hormone replacement therapy in primary care*. British Medical Journal. 2001;323(7311):490-3.