

Multiple criteria decision analysis for biomarker prioritization: Developing a socio-technical approach to assist researchers and clinicians in biomarker selection for validation and translation into clinical applications

Beatriz Norte

Instituto Superior Técnico - Universidade de Lisboa

Abstract—Although thousands of biomarkers have already been discovered, not many have been applied into clinical practice, due to the high time and monetary resources necessary for their identification and validation. Due to its complexity and heterogeneity, and consequent lack of specific biomarkers, COPD is ideal to apply the model developed in this thesis. With the purpose of evaluating and selecting the most promising COPD prognostic biomarkers, a MCDA model was developed, with focus on the structuring phase, using a socio-technical approach based on the MACBETH method, including an assessment of areas of concern in the biomarker field, the definition of criteria for biomarker selection (using literature, interviews and a Web-Delphi) and the design of the resultant value tree. Results show that experts believe that the clinical relevance, clinical added value, quality of studies and test reliability are relevant criteria, while the costs of development, the patient comfort and the easiness to measure, analyse and interpret are secondary, as one must give up certain benefits if the biomarker significantly improves the patients' health/well-being. Although ten evaluation dimensions were considered relevant by experts, some dependencies were found, leading to the grouping and reformulation of the dependent dimensions. Despite some difficulties, the approach applied in this thesis worked well, resulting in a good structure for the MCDA model, with seven well defined and relevant evaluation criteria for the prioritization of biomarkers. In the future, it would be interesting to complete the model, including the building, testing and validation phases of the model.

I. INTRODUCTION

According to the National Institute of Health Biomarkers Definitions Working Group, a biomarker is a “characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention” [1]. Recently, as a result of the advancement in personalized medicine, a biomarker can also be perceived as an indicator that enables the development of treatment interventions for specific patients, maximizing therapeutic benefits and minimizing the risk of treatment [2].

Biomarkers are used in all stages of drug development, being a valuable tool to create safe and efficacious drugs, by providing greater accuracy or more complete information regarding disease progression or drug performance [3]. They

can be classified as one, or several, of the following categories, depending on their use and clinical context: diagnostic, prognostic, predictive, monitoring, response, safety or susceptibility biomarker [4].

Thousands of proteins have already been proven to be hallmarks of emerging disease, prognosis of a patient or response to treatment. The identification of protein biomarkers associated with a disease is essential for the improvement of personalized medicine based on blood tests, since they have a great impact in drug discovery and development, possibly leading to better disease outcomes, such as higher patient survival and lower healthcare costs, among others. However, even though several biomarkers have been discovered and presented in studies, there are not many that have been applied into clinical practice (only about one hundred and fifty out of thousands of identified biomarkers), due to the fact that the process of identification and validation of disease specific biomarkers is very time consuming and requires many resources, as well as to the fact that some studies that determine the clinical value of a biomarker are not reproducible and that some studies do not match regarding requirements for regulatory and marker approval [5]. In fact, the whole process of biomarker discovery and development is estimated to cost between 800 million and 1.7 billion dollars, requiring between 7 and 12 years [3][4]. Therefore, it is necessary to narrow down the number of candidate biomarkers, in order to start the validation and translation into clinical applications process with the best possible candidates, so the investment, both in time and money, can be minimal, while achieving the best possible results.

In order to increase the number of clinically validated biomarkers, instead of increasing the number of studies that discover new ones, the Cost Action European project, Clin-iMark, aims to improve the quality and reproducibility of studies and to create the Best Biomarker Practice guidelines, in order to establish a coherent biomarker development pipeline from discovery to clinical application, which shall provide guidance to [5]:

- Classify biomarkers considering their characteristics, pre-

- dicted clinical use and phase of development;
- Select and validate appropriate research-grade biomarker detection tests;
- Select studies and biological samples that have been designed appropriately and that can be reproduced, to reliably validate biomarkers on a clinical level;
- Select and report on appropriate clinical data storage, biomarker data storage, data analysis protocols, privacy concerns, ethical issues, and statistical analysis methods.

To demonstrate this project, which is also being developed with the MEDI-VALUE research project, which develops collaborative approaches in Multiple-Criteria decision analysis (MCDA) to evaluate medical devices, Chronic Obstructive Pulmonary Disease (COPD) will be used as an example. COPD is a complex and heterogeneous disease, caused mainly by cigarette smoke, that is characterized by persistent and progressive airflow limitation and associated with an abnormal chronic inflammatory response of the lungs due to noxious particles and gases, leading to an acceleration of structural changes and narrowing of airways and, consequently, decline in lung function [6][7]. It has been reported that COPD kills more than four million people per year [8] and that by association with its high prevalence, COPD generates significant social and healthcare costs [7], resulting in \$44 billion healthcare expenses every year [8]. Due to its complexity and heterogeneity, as well as to its molecular and clinical characteristics, there is great difficulty to efficiently stratify the patients and to introduce personalized therapeutic approaches, which, together with the fact that the currently available clinical tools do not predict the progression and exacerbations of this disease efficiently [5], makes COPD a great candidate to demonstrate this project with, due to its clear need for new drugs that can solve some of its major clinical problems: and it all starts with biomarker prioritization.

With all this in mind, an extensive literature survey about COPD biomarkers has been conducted, considering all literature published between 2016 and 2018, plus the biomarkers and articles mentioned in S. Ongay et al. [9], resulting in a list of about one hundred of the most promising candidate biomarkers for COPD. There are several proteins in this list that showed interesting measurement values, with distinct statistical power, as well as interesting outcomes, with high relevance to COPD, when investigated by different research groups. Some proteins had different applications suggested by different authors, likely due to the objective of each study, and there were several proteins that showed potential for more than one outcome, such as diagnostic and prognostic, or context of use. The same way, one specific prognosis, such as mortality or exacerbation, can have several biomarkers associated with distinct biological functions. However, it is very likely that not all selected biomarkers will be relevant in drug development in COPD.

For this reason, it is now necessary to evaluate all the biomarker candidates and select the most promising ones, to be further analysed and tested. MCDA methods and tools have been identified as a good option to reach this goal, by

helping researchers in optimal decision making and involving the knowledge of multiple stakeholders (researchers, clinicians and pharmaceutical industry), while building an evaluation model. Therefore, the proposed solution to the problem above, and main goal of this thesis, is to create and test a MCDA approach, with focus on the structuring of the model, to help researchers and clinicians in COPD biomarker prioritization, but that can be adapted and used for any given disease, and not only COPD, which will be used solely as an example to apply the created model and to demonstrate and evaluate its results. Furthermore, by facilitating the biomarker prioritization process, and by selecting the most promising biomarkers, this approach will help reducing significantly both time and money resources associated with the process of identification and validation of disease specific biomarkers.

II. LITERATURE REVIEW

A. Prioritization Approaches - Multiple Criteria Decision Analysis and the MACBETH approach

There are several decision making techniques available, but we chose to use MCDA models for biomarker prioritization in this thesis, because they are not only some of the most commonly used approaches for priority-setting [10], but they can handle problems with multiple objectives, have an encompassing nature, are intuitive, have theoretically sound methods to balance benefits, costs and risks, allow for the use of different types of data, allow the involvement of different stakeholders, taking their preferences and values under consideration, have easily understandable outputs [11][12], and promote transparency, accountability and reasonableness in decision-making [13]. This technique also has the great advantage of capturing the knowledge from a decision, making it reusable for others who need to make the same decision, or a similar one [14]. Considering that the purpose of this thesis is to create a generic socio-technical approach to assist researchers and clinicians in the selection of the most promising biomarkers for validation and translation into clinical applications, the previously mentioned advantages of MCDA make it a very promising approach.

MCDA is a tool used for decision making, that uses a set of quantitative and qualitative approaches, that simultaneously and explicitly takes under consideration multidimensional and often conflicting factors, allowing for comparison of technologies, namely medical, by combining individual criteria into one overall assessment. This approach has the potential to overcome the challenges presented by traditional decision-making tools, in particular when the decision-making is complex and includes multiple criteria, multiple stakeholders and both quantitative and qualitative data [15], which is ideal in Healthcare Technology Assessment (HTA), where decisions are not only complex, but also involve uncertainties and have to consider the preferences and values of stakeholders.

According to A. Angelis and P. Kanavos [16], there are several MCDA approaches, which can be classified in three different groups: (a) value measurement methods, which includes Multi-Attribute Value Theory (MAVT), (b) outranking

methods and (c) ‘satisficing’ and aspiration level methods. In this thesis, we wanted to use a value measurement method, due to the simplicity of the required value judgements and to the fact that they can be applied to multiple decision contexts. More specifically, we wanted to use a MAVT method, since they are comprehensive, robust and they are able to reduce motivational biases and ambiguity [16]. As a result, AHP, as well as all the other MCDA methods that do not respect the theoretical foundations of MAVT, has not been considered to be used in this thesis, even though it is one of the most frequently used and implemented MCDA methods [17].

Therefore, we analysed some of the most well known and most used MCDA methods that respect the theoretical foundations of MAVT, according to C. Bana e Costa [18], including bisection techniques and direct rating (methods used for scoring), trade-off, swing weighting and point allocation (methods used for weighting) and MACBETH (method used for both scoring and weighting), in order to select the most suitable one for the objective of this thesis.

After analysing all the options, while considering the characteristics of the problem presented in this thesis, we opted to use Measuring Attractiveness by a Categorical-based Evaluation Technique (MACBETH) as our multiple criteria decision method, since it can be used for both scoring and weighting, it is a non-numerical approach, which makes it easier for the decision-maker to express value judgements, and because we considered it the best option to answer the problem presented.

According to C. Bana e Costa et al. [19], MACBETH is a “humanistic, interactive and constructive approach to the problem of how to build a quantitative model of values based on qualitative (verbal) difference judgements, that facilitates the path from ordinal to cardinal preference modelling, namely analysing judgmental inconsistency and offering suggestions to move the process forward”. The fact that this method allows quantitative models to be built based on qualitative judgements makes the whole process significantly simpler, which is a great advantage.

The MACBETH process, like any multicriteria analysis process, usually starts with an analysis of the context of the problem, followed by a discussion and definition of the exclusion and evaluation criteria, as well as the descriptors of performance (either quantitative or qualitative). The next step is the definition of the levels of performance and reference levels to be used for building the value function. Next, the MACBETH protocol for questioning the decision-makers is used, in order to obtain pairwise absolute judgements on differences of attractiveness between levels of performance or options of the criterion, using the semantic scale: null, very weak, weak, moderate, strong, very strong, extreme. If the answers are consistent, the M-MACBETH software will then suggest a quantitative value scale, through the resolution of a linear programming problem, which should be adjusted and validated with the decision-maker. In case the answers are not consistent, M-MACBETH will suggest an alternative to make them consistent. Following the creation of the value scale, the global value of each proposition is calculated, using

the additive aggregation model. Finally, a robustness and sensitivity analysis of all the options considered to answer the problem in hand is performed. With all the information obtained, one of the options will be recommended as the best or most promising [19][20].

Although there are several examples in literature about the use of MCDA and MACBETH in healthcare, namely for supporting planning of decisions in the long-term care sector [21], hospital auditing [22], building a population health index [23], assisting in the diagnosis of Alzheimer’s disease [24] and evaluating health and safety risks [25], there are very few studies regarding the use of MCDA in biomarker prioritization, one example being the study by A. Miquel-Cases et al. [26], where MCDA is used, among other techniques, to select the most promising biomarkers for breast cancer, but with no particular focus on it.

As for the specific case of COPD, although there are articles where MCDA is used in association with it, namely in the study by K. Marsh et al. [27], where an evaluation of COPD treatments is made using MCDA, there is almost no information regarding biomarker prioritization in COPD using this method, even though there are some studies where biomarkers are prioritized in some other way, namely in the study by S. Ongay et al. [9], where biomarkers are prioritized by applying several criteria individually, including the number of individuals considered in the cohorts and statistically significant differences between COPD patients and healthy smokers.

B. Participatory Methods

As MACBETH includes both technical and social components, it is a socio-technical approach [28]. Consequently, it requires the appropriate usage of participatory methods.

A participatory method is an approach that involves “the public” relevant to the topic being evaluated in decision-making processes. The public can include citizens, stakeholders of a particular project, experts or even members of private industry and government. When applied early in a decision-making process, participatory methods allow the participants to share their perspectives on the issue being evaluated, being useful to achieve consensus when there are differences in opinion or conflicts among the participants. Using these methods allows all voices to be heard, leading to a more democratic decision-making process and to an improvement of the quality of the decisions [29].

However, there are many participatory methods available. Some of the better known and most used ones, based on N. Slocum [29], are: charrette, citizens jury, decision conference, delphi, expert panel, interviews, participatory assessment, monitoring and evaluation, planning cells, scenarios workshop and world café. When deciding which method to use, it is important to consider five elements: objectives, nature and scope of the issue, participants, time available and budget [29].

During the structuring of the model, which is the focus of this thesis, we want to interview a few experts, followed by the gathering of the opinion of a great number of experts from different geographies in an easy and confidential way

(therefore, an online method would be preferable), the time available is very reduced and there is no budget. Therefore, the interviews and the Delphi method are the methods that will be used as they are the ones that responded better to the presented specifications. During the remaining phases, the Delphi method would be used as well, but we would also want to confirm our results and decisions, in order to reach consensus, and to validate the model, for which a decision conference would be the ideal choice. Therefore, a collaborative value model, which is a socio-technical model that combines multicriteria decision conferencing and the Delphi participatory process, with the goal of building widely informed evaluation models, will be applied in this thesis. This model allows for a unique collaborative learning, where participants are seen not only as experts providing individual expertise, but also as collective learners that construct shared judgemental knowledge, leading to widely informed and more acceptable evaluation models [28].

III. METHODOLOGY

In this thesis, and with the purpose of designing an approach that allows for the prioritization of COPD prognostic biomarkers and, therefore, the reduction of time and monetary resources associated with the process of identification and validation of disease specific biomarkers, a socio-technical approach was used, which includes the technical elements of a MCDA model, but also the social elements of participatory methods, resulting in an approach where scientific evidence is combined with the opinion of different stakeholders and experts, where there is better communication and a shared understanding of the issues and where a sense of common purpose is generated [11][30]. By using multiple sources of information (researchers and clinicians), the amount of information available to build the model is maximized, the potential impact of sources that rely on unreliable and inaccurate information is reduced and the model becomes more inclusive and representative, leading to more validity and credibility [30].

As demonstrated in Figure 1, the MACBETH socio-technical approach to build a multicriteria model to prioritize biomarkers can be divided in three main phases: model structuring, model building and model testing and validation. These phases are then divided in technical and social approach: the technical component involves the use of MACBETH, while the social approach has the purpose of providing the necessary information to build the multicriteria model, using web-based Delphis, interviews and decision conferences, because it was decided that the model created in this thesis would be based on collaborative value modelling, since it elicits and analysis individual judgment knowledge from a very large and diverse group of individuals [28]. However, in this thesis the focus was on the structuring phase.

First, it was necessary to have a facilitation team, which was responsible for managing the dynamics of the whole process and for guaranteeing its success, as well as for conducting the activities during MCDA model building [28].

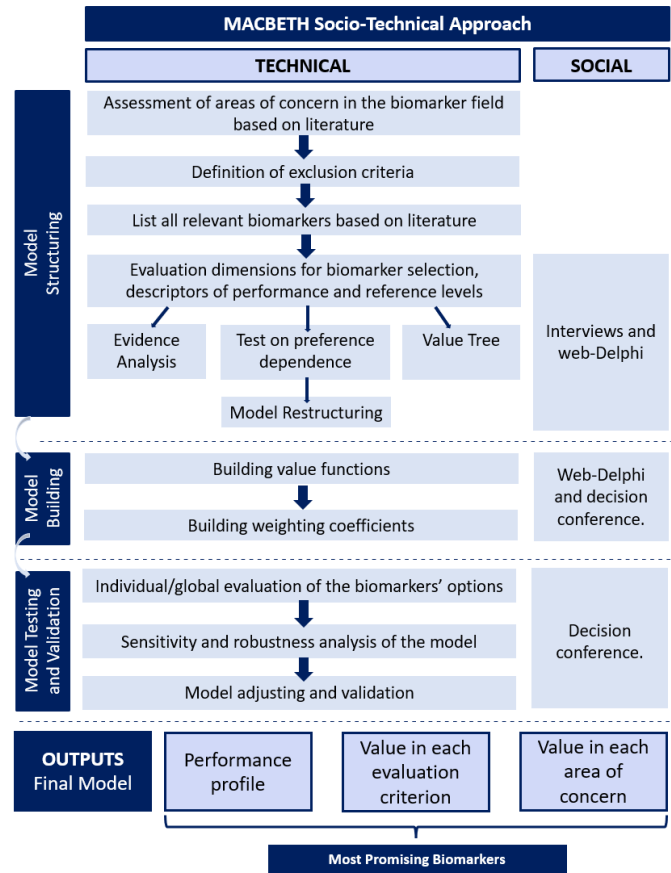


Figure 1: MACBETH socio-technical approach design for the prioritization of biomarkers (inspired in the methodology of the EURO-HEALTHY project [31] and adapted to the biomarker context).

In the model structuring phase, we started by assessing the areas of concern in the biomarker field, based on literature and on the input from the decision-maker, in order to understand what was the problem, the key components of biomarkers' benefits, risks and costs and what needed to be done. After the areas of concern had been defined, the next step was the definition of exclusion criteria, which would allow for the automatic rejection of some of the biomarker options. This was followed by the creation of a list that included all relevant biomarkers' options to be considered in the decision-making, based on literature. After the list was complete with all the biomarkers approved after the application of the exclusion criteria, the evaluation criteria to be used in the decision-making process were determined, along with their descriptors of performance and reference levels, which should address the objectives and concerns that experts find fundamental to evaluate the biomarker options. Before the evaluation criteria are confirmed to be preference independent, they are called evaluation dimensions. Therefore, first of all, a list of possible EDs was created based on evidence analysis, and the correspondent descriptors of performance and reference levels *good* and *neutral* were defined. Secondly, this list was

presented to a few selected experts and a semi-structured interview was conducted with the purpose of changing, adding or eliminating any ED. Finally, in the form of a web-based Delphi, a larger group of experts was presented with the updated list, where they had to give individual qualitative judgements regarding the relevance of each proposed ED, but could also add comments, in order to reach an agreement about the EDs to be used. After this, a test on preference dependence was conducted (where two experts were interviewed) to verify if all EDs were independent, followed by a restructuring of the model due to some dependencies that were found, resulting in a final list of independent EDs, called evaluation criteria. The chosen criteria were then assigned to one area of impact (benefits, costs or risks) and, finally, a value tree was designed, with the criteria divided according to their area of impact.

IV. RESULTS

A. Areas of Concern and Context of Use

Based on evidence analysis and with input from the decision-maker, it became clear that the most relevant areas of concern in the biomarker field are benefits for the patient, the associated costs and the eventual risks. Therefore, each evaluation criteria to be defined will be included in one of these three categories.

Among the seven different biomarker categories, there are different characteristics and objectives. Consequently, comparing biomarkers that belong to different categories is substantially more complex than comparing biomarkers among the same category. Therefore, in order to reduce the complexity of the model, we opted to focus in only one biomarker category and context of use (COU), reducing, as a consequence, the number of candidate biomarkers. The choice was based on the biomarker category and respective COU that presented the largest number of biomarkers investigated, reflecting the greatest clinical needs in COPD, based on evidence from literature. Biomarkers for prognosis of exacerbation and mortality are the most investigated category of biomarkers until this day. Therefore, we will be focusing on this biomarker category and its associated COU will be to enrich clinical trial/further validation research for an event or population of interest.

B. Definition of Exclusion Criteria

Before defining the evaluation criteria, it is necessary to define the exclusion (or screening) criteria, which are used to automatically reject some biomarkers.

These criteria are established by the decision-maker and, for biomarker prioritization, they are: biomarkers cannot be validated or in the qualification phase by Food and Drug Administration (FDA); biomarkers cannot have a low amount of information associated; and biomarkers must be associated with studies that analyse more than one hundred individuals.

Only the biomarkers that are approved after the application of the exclusion criteria are considered in the model.

C. List of Biomarker Options

When this thesis was proposed, there was already a list of over one hundred biomarkers to be considered, obtained from literature published between 2016 and 2018. As a complement, and based on the work of S. Ongay et al. [9], new biomarkers found in literature until 2016 were added to the initial list.

After all possible biomarker options were gathered, they were divided according to their category. Since we chose to work exclusively with prognostic biomarkers, only those were selected. Afterwards, the remaining options were filtered using the exclusion criteria previously defined.

D. Definition of Evaluation Dimensions, Descriptors of Performance and Reference Levels

D.1. Evidence Analysis

This phase consisted on the definition of a list of evaluation dimensions (EDs), which are evaluation criteria that have not been proven to be independent yet, based on literature, and subsequent formulation of their descriptors of performance and reference levels based on logic and examples and information found in literature. As a result, an initial list of nine evaluation dimensions and their descriptors of performance and reference levels were defined. The nine evaluation dimensions were: clinical added value, clinical relevance, patient comfort, quality of evidence, easiness to measure, analyse and interpret, reproducibility of results, expected utility in drug development, costs of validation and utility and ethical issues.

D.2. Interviews

In this phase, four experts (two researchers and two clinicians, members of CliniMark) were interviewed via video call, with the purpose of analysing the relevance of each of the previously defined evaluation dimensions, verifying the quality of their descriptions and descriptors of performance and the rightness of the reference levels attributed, suggesting alterations, if necessary, and suggesting new EDs believed to be relevant for the purpose at hand.

In a general way, the four interviewed experts agreed with the EDs presented, their descriptions and respective descriptors of performance. However, there were several suggestions made and, in some cases, some differences in opinion between the experts. As a result, a new ED was created and several EDs were reformulated (either names, descriptions, descriptors or reference levels). For example, the quality of evidence became quality of the study and a new ED was created, level of evidence, as a complement to it. Also, the reproducibility of results became test reliability because, if the reliability of a test is high, then that test is accurate, reproducible and consistent from one trial to another. This way, the ED became more complete. As for the costs of validation and utility, it was suggested that the ED name should be "Costs of Development", as the development phase includes both the analytical validation and the utility studies. In the end, a list of ten evaluation dimensions was obtained: clinical added value, clinical relevance, patient comfort, level of evidence, quality of the study, easiness to measure, analyse and interpret, test

reliability, potential value to address an unmet need in drug development, costs of validation and utility and ethical issues.

D.3. Web-based Delphi

Contrarily to the two previous phases, this one was designed with the sole purpose of determining the relevance of the previously defined evaluation dimensions, based on the opinion of a great number of experts, instead of defining and improving the EDs, descriptors of performance and reference levels.

In this phase, regardless of the round, the 22 experts (researchers and clinicians, all members of CliniMark) that participated in the three rounds were asked to evaluate the relevance of the ED obtained after the interviews, which were inserted in the web-based Delphi as indicators, each with a description associated containing not only the description of the ED in question, but also the descriptors of performance and reference levels associated so the experts participating could inform themselves before deciding on its relevance.

Being able to change or maintain their opinions after analysing the answers of the other participants, led to an increased agreement between participants throughout the three rounds. Three rounds proved to be enough to reach stability in responses and consensus between the participants.

There was a solid number of participants throughout the three phases, with only three drop-outs out of twenty-four participants (13% drop-out rate), resulting in an inclusive and representative and, consequently, valid and credible model. The rigor was maintained throughout the whole web-Delphi process, as the response rate surpassed 90% in both round two and three. During the whole Web-Delphi process, reminders were sent regularly to experts who had not completed the questionnaire yet and deadlines were extended in the first two rounds. This two measures helped increasing the number of participants during the first round and decreasing the drop-out rate in the second and third rounds, leading to a greater representativeness of all points of view.

After the third round, EDs were approved either if the percentage of *Strongly Agree* votes was superior to 50% and the percentage of *Disagree* and *Strongly Disagree* votes was inferior to 33% or if the percentage of *Strongly Agree* votes, plus the percentage of *Agree* votes, was superior to 75%. As we can observe in Figure 2, where the results of the third round are presented, all ten EDs were considered to be relevant by more than 75% of the experts.

As one cannot assume that the results of the Web-based Delphi are one hundred percent trustworthy, since the participants do not interact with each other and no numerical judgements are obtained, the results were presented to Doctor Deborah Penque, proteomics expert, who did not participate in the process, so she could decide if, based on the results and in her expert opinion, every evaluation dimension, specially the ones with lower agreement percentage, should actually be considered as relevant. In her opinion, since we are in an exploratory phase, we should accept all evaluation dimensions, including those that raised some doubt, because even those presented a high percentage of acceptance. As a result, the ten

Indicador	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
1 - Clinical Added Value	88%	14%			
2 - Clinical Relevance	95%	5%			
3 - Patient Comfort	24%	52%	19%	5%	
4 - Level of Evidence	78%	24%			
5 - Quality of the Study	86%	14%			
6 - Easiness to Measure, Analyse and Interpret	29%	52%	14%	5%	
7 - Test Reliability	88%	14%			
8 - Potential Value to Address an Unmet Need in Drug Development	19%	71%	10%		
9 - Costs of Development	5%	76%	14%	5%	
10 - Ethical Issues	33%	57%	10%		

Figure 2: Results of the third and last round of the Web-Delphi to determine the relevant evaluation dimensions for biomarker prioritization.

proposed evaluation dimensions were considered to be relevant for the prioritization of prognostic biomarkers in COPD.

D.4. Test on Preference Dependence

This phase started with a reflection regarding the possibility of some evaluations dimensions depending on each other.

To facilitate the process, two clusters were created: one regarding evaluation dimensions related to the clinical benefit of the biomarkers (which included clinical added value, clinical benefit and potential value to address an unmet need in drug development), and another related to the the studies and methodologies associated with each biomarker (which included level of evidence, quality of the study, test reliability and easiness to measure, analyse and interpret).

After pondering on the possibility of eventual dependencies among this evaluation dimensions, the following groups were considered to be possibly preference dependent:

- Clinical added value, clinical benefit and potential value, because these three EDs relate to the clinical benefit of the biomarker. Therefore, preference dependence must be tested for the three possible pairs.
- Quality of the study and level of evidence, because these EDs are both connected with the study.
- Quality of the study and test reliability, because for a study to be good and trustworthy, the tests performed must be reliable.

The easiness to measure, analyze and interpret was not considered to be possibly preference dependent on another ED, namely the test reliability or the quality of the study, because it has no direct connection with the quality or reliability of the study, but with the easiness to apply the methodology and analyze the results. The remaining EDs were not considered to have any connection in terms of preference with the other EDs.

After the evaluation dimensions to be tested were defined, two experts in the field of biomarkers were interviewed via video call, with the purpose of determining if there were preference dependencies among each pair. The results of the test can be observed in Table I.

Considering the results obtained, where three pairs of evaluation dimensions were found to be preference dependent.

Table I: Results of the test on preference dependence for each of the four swings (S_x): unilateral preference dependence (UD), bilateral preference dependence (BD) or preference independence (I). The judgements were made based on the MACBETH semantic scale: null (N), very weak (VW), weak (W), moderate (M), strong (S), very strong (VS), extreme (E).

Evaluation Dimensions	S_1	S_2	S_3	S_4	Result
Clinical Relevance/ Clinical Added Value	S/ VS	S/ VS	W	M	UD
Clinical Relevance/ Potential Value to Address an Unmet Need in Drug Development	W	W	VS	VS	I
Clinical Added Value/ Potential Value to Address an Unmet Need in Drug Development	W	W	S	VS	I
Level of Evidence/ Quality of the Study	M/ S	W	M	M	UD
Quality of the Study/ Test Reliability	M	W	M/ S	M	BD

dent (either uni or bilaterally), it was necessary to redefine them by combining them into single independent evaluation dimensions. As a result, the five not preference dependent evaluation dimensions and the two newly created evaluation dimensions, may now be called evaluation criteria, as they are independent. The two new criteria resultant from the five preference dependent ones, including definition, descriptors of performance and reference levels, were approved by the two experts who participated in the interview to assess preference dependence. The definition of the seven resultant evaluation criteria can be observed in Table II, and their respective descriptors of performance and reference levels in Table III. As a result, the value tree presented in Figure 3, where the criteria are divided in areas of concern and clusters, was built.

V. DISCUSSION/CONCLUSION

Although MCDA, and most specifically the MACBETH approach, has been applied in several health-care related studies, there are very few studies regarding its use in biomarker prioritization, and almost no information regarding biomarker prioritization in general in COPD. Therefore, by applying the MACBETH socio-technical approach to the case of COPD in this thesis, a new and original model has been designed and partially developed (only the structuring phase), for the purpose of evaluating and selecting the most promising COPD prognostic biomarkers among the hundreds found in literature.

Several steps were taken to reach a final list of valid and credible criteria. The evidence analysis allowed for an initial and very basic list of nine potentially relevant evaluation dimensions. The interviews were essential to reach a group

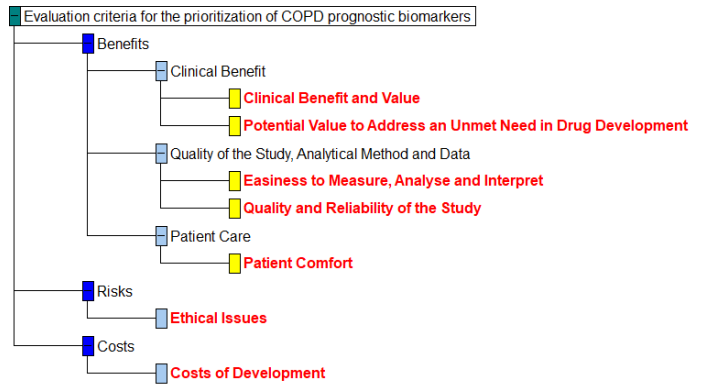


Figure 3: Organization of the criteria in a value tree using the software M-MACBETH. The dark blue nodes are the areas of concern, the light blue nodes with the terms not highlighted in red are the clusters and the evaluation dimensions are the terms highlighted in red.

of ten well-defined, reformulated and potentially relevant biomarkers and respective descriptors of performance and reference levels. The twenty-two expert opinions gathered during the web-based Delphi were essential to confirm the relevance of the proposed evaluation dimensions. The opinions clearly converged throughout the three rounds, meaning the Delphi process was successful in promoting agreement. Finally, some dependencies were found during the test on preference dependence, which led to the grouping of some evaluation dimensions, resulting in a total of seven preference independent evaluation dimensions (when independent, called evaluation criteria).

It is particularly relevant to mention that most literature on MCDA in HTA does not test or treat evaluation dimensions for inter-dependencies. Therefore, by testing for inter-dependencies in this thesis, we were able to guarantee that the final list of relevant evaluation criteria were independent of each other, which is an essential characteristics for criteria in a MCDA model. Furthermore, by including experts from different groups of expertise and geographical locations throughout the whole process, the model became more inclusive and representative, leading to more validity and credibility.

The main difficulties throughout the whole process involved selecting the criteria and building clear and non-subjective descriptors of performance, motivating the experts to participate in the Web-based Delphi, explaining to the experts what was being asked and what was the purpose of the test on preference dependence and the fact that a better development of methods would require more resources than the ones that were available.

Overall, the socio-technical methodology used in this thesis worked well, despite the difficulties met throughout the process, resulting in a good structure for the MCDA model, with seven well defined and relevant evaluation criteria that will be of great help for CliniMark to prioritize COPD prognostic biomarkers, by facilitating the comparison of the different

Table II: Final list of evaluation criteria for biomarker prioritization and respective descriptions.

Evaluation Dimensions	Description
Clinical Benefit and Value	Represents how relevant the biomarker is in clinical practice, in terms of its ability to improve patient well-being and outcome and how much value it adds (if it brings something new/different) in relation to the already qualified biomarkers, based on the AUC between the new biomarker and the already qualified ones.
Patient Comfort	Comfort of the patient, considering how invasive the procedure used to access the biomarker is (if it is accessible in the peripheral tissue, namely in blood or urine, the discomfort will be null/minimal; if not the discomfort may be greater.
Quality and Reliability of the Study	A high quality and reliability of a study is based on: i) high methodological quality, validity and applicability to patient care; ii) well-designed cohort, including adequate biospecimen repositories and sample size; iii) proper analytical and statistical methods; iv) consistent and reproducible results across time (test-retest reliability), across items (internal consistency) and across researchers (inter-rate reliability); v) accurate data analysis and interpretation to enable biomarker identification/validation meeting pre-specified performance criteria for a given clinical application context
Easiness to Measure, Analyse and Interpret	Easiness to measure and analyze the biomarker, considering the simplicity of the method used, the time needed for this purpose and the need for high-throughput, as well as the easiness to interpret the results.
Potential Value to Address an Unmet Need in Drug Development	Potential value of a biomarker for a particular context of use (COU) in drug development, likely to be recognized by the FDA as qualified biomarker in the near future. COU in drug development includes patient stratification, selection, trial enrichment, dose selection, response/efficacy assessments, safety assessments, among others.
Costs of Validation and Utility	Amount of money necessary for the validation of a biomarker (where the generalizability across different samples and the reproducibility and standardization of the assay are determined) and utility studies (where the results must show performance characteristics, well-designed experiments and the added value in research models and/or patients).
Ethical Issues	Ethical issues associated with the application of the biomarker in clinical practice, including psychological reactions (associated with factors such as risk of catastrophic reaction, no proven long-term treatments and risk of false positive) and social stigma.

options and by reducing time and monetary resources.

In the future, when the model is completed, it will be possible to select the most promising COPD biomarkers, which will increase significantly the probability of the biomarkers' being validated and qualified successfully. As a consequence, personalized medicine can be improved, by improving disease specific drugs using the qualified disease specific biomarkers, which will lead to an increase of the quality and efficiency of patient care and, possibly, to better disease outcomes, such as higher patient survival and lower health costs, among others. Therefore, the model designed and developed in this thesis comes as a great contribute in the drug development and clinical fields, as it is innovative and allows for the determination of the most promising disease specific biomarkers, when there are hundreds available, leading to the development of more efficient drugs and, consequently, better patient care.

VI. FUTURE WORK

In the future, it would be interesting to complete the model for COPD prognostic biomarkers, including the building, testing and validation of the model.

It would also be interesting to apply the developed MCDA model to a COU different from the one used in this thesis. It would be ideal to develop a new MCDA model that could compare and prioritize biomarkers with several different COUs, instead of the the single one considered during this thesis, for the purpose of simplicity. This would result in a more comprehensive model.

It would also be interesting to improve MCDA tools for stakeholder involvement and evaluation in the context and to develop a model with the participation of different combinations of participants, as well as a greater number of experts, from several fields, so that the model can be more robust, avoiding the variability associated with a group of decision-makers as small as the one used in this thesis.

Finally, the developed model could be applied to other biomarkers and, specially, diseases other than COPD, since the idea of developing such a model was to make it comprehensive, in such a way that it was not limited to the prioritization of biomarkers for only one disease.

Table III: Final list of evaluation criteria and respective descriptors of performance and reference levels *neutral* and *good*.

Evaluation Dimensions	Descriptors of Performance
Clinical Benefit and Value	<ul style="list-style-type: none"> • Very relevant, showing ability to improve well-being and outcome to the generality of patients and adding over 10% value to the already existing biomarkers. [GOOD] • Very relevant, showing ability to improve well-being and outcome to the generality of patients and adding 10% or less value to the already existing biomarkers. • Relevant, showing ability to improve well-being and outcome to patients with specific characteristics and adding 10% or less value to the already existing biomarkers. • Relevant, showing ability to improve well-being and outcome to patients with specific characteristics but adding no value to the already existing biomarkers. [NEUTRAL] • Irrelevant, showing poor or no ability to improve patient’s well-being and outcome and adding no value to the already existing biomarkers.
Patient Comfort	<ul style="list-style-type: none"> • Non-invasive procedure, with no discomfort to the patient. [GOOD] • Semi-invasive procedure, with mild discomfort to the patient. [NEUTRAL] • Invasive procedure, with mild/moderate discomfort to the patient. • Invasive procedure, with severe discomfort to the patient.
Quality and Reliability of the Study	<ul style="list-style-type: none"> • High Quality: all recommendations are fully accomplished. [GOOD] • Medium Quality: three/four recommendations are fully accomplished. • Low Quality: one/two recommendations are fully accomplished. [NEUTRAL] • Very low quality: no recommendation is accomplished.
Easiness to Measure, Analyse and Interpret	<ul style="list-style-type: none"> • Easy to measure and analyze the biomarker and easy to interpret the results. [GOOD] • Hard to measure and analyze the biomarker and easy to interpret the results. • Easy to measure and analyze the biomarker and hard to interpret the results. [NEUTRAL] • Hard to measure and analyze the biomarker and hard to interpret the results.
Potential Value to Address an Unmet Need in Drug Development	<ul style="list-style-type: none"> • Potential high/moderate-value for a particular COU in drug development with high unmet need. [GOOD] • Potential high/moderate-value for a particular COU in drug development. [NEUTRAL] • Potential low-value for a particular COU in drug development.
Costs of Development	<ul style="list-style-type: none"> • 5M\$ [GOOD] • 10M\$ • 15M\$ • 20M\$ [NEUTRAL] • 25M\$ • 30M\$
Ethical Issues	<ul style="list-style-type: none"> • Presents no risk of psychological reactions and/or social stigma to the patient. [GOOD] • Presents low risk of psychological reactions and/or social stigma to the patient. • Presents moderate risk of psychological reactions and/or social stigma to the patient. [NEUTRAL] • Presents high risk of psychological reactions and/or social stigma to the patient.

REFERENCES

- [1] K. Strimbu and J. A. Tavel, “Biomarkers In Risk Assessment: Validity And Validation,” *Environmental Health*, vol. 5, no. 6, p. 144, 2001.
- [2] Z. Hollander, M. L. DeMarco, D. D. Sin, M. Sadatsafavi, R. T. Ng, and B. M. McManus, “Biomarker Development in COPD,” *Chest*, vol. 151, no. 2, pp. 455–467, 2016.
- [3] M. R. Bleavins, C. Carini, M. J. Romet, and R. Rahbari, “Biomarkers,” *Pharmaceutical Sciences Encyclopedia*, 2010.
- [4] C. V. Gerlach, M. Derzi, S. K. Ramaiah, and V. S. Vaidya, “Industry Perspective on Biomarker Development and Qualification,” *Clinical Pharmacology and Therapeutics*, vol. 103, no. 1, pp. 27–31, 2018.
- [5] “CA16113 - CliniMARK: ‘good biomarker practice’ to increase the number of clinically validated biomarkers.” European Cooperation in Science and Technology. Available at: <https://www.cost.eu/actions/CA16113/#tabs—Name:overview> (Accessed: 21-03-2019).
- [6] T. Aggarwal, R. Wadhwa, N. Thapliyal, K. Sharma, V. Rani, and P. K. Maurya, “Oxidative, inflammatory, genetic, and epigenetic biomarkers associated with chronic obstructive pulmonary disorder,” *Journal of Cellular Physiology*, vol. 234, no. 3, pp. 2067–2082, 2019.
- [7] J. Gea et al., “The BIOMEPOC Project: Personalized Biomarkers and Clinical Profiles in Chronic Obstructive Pulmonary Disease,” *Archivos de Bronconeumologia*, no. xx, 2018.
- [8] E. Aydinoglu, D. Penque, and J. Zoidakis, “Systematic review on recent potential biomarkers of chronic obstructive pulmonary disease,” *Taylor & Francis Online*, 2018.

- [9] S. Ongay, F. Klont, P. Horvatovich, R. Bischoff, and N. H. Ten Hacken, "Prioritization of COPD protein biomarkers, based on a systematic study of the literature," *Advances in Precision Medicine*, vol. 1, no. 1, p. 4, 2016.
- [10] P. K. Narayan and S. Narayan, "Using economic evidence to set healthcare priorities in low-income and lower-middle-income countries: a systematic review of methodological frameworks," vol. 1186, no. December 2007, pp. 1171–1186, 2008.
- [11] M. D. Oliveira, I. Mataloto, and P. Kanavos, "Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art," *The European Journal of Health Economics*, no. 0123456789, 2019.
- [12] A. D. Montis and P. D. Toro, "Assessing the quality of different MCDA methods," *Alternatives for environmental valuation*, no. June 2016, pp. 99–184, 2000.
- [13] R. Baltussen, M. Paul Maria Jansen, L. Bijlmakers, J. Grutters, A. Kluytmans, R. P. Reuzel, M. Tummers, and G. J. v. der Wilt, "Value Assessment Frameworks for HTA Agencies: The Organization of Evidence-Informed Deliberative Processes," *Value in Health*, vol. 20, no. 2, pp. 256–260, 2017.
- [14] "Decision Making Techniques." Decision Innovation. Available at: https://www.decision-making-solutions.com/decision_making_techniques.html (Accessed: 17-04-2019).
- [15] G. Adunlin, V. Diaby, and H. Xiao, "Application of multicriteria decision analysis in health care: A systematic review and bibliometric analysis," *Health Expectations*, vol. 18, no. 6, pp. 1894–1905, 2015.
- [16] A. Angelis and P. Kanavos, "Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework," *Social Science and Medicine*, vol. 188, pp. 137–156, 2017.
- [17] M. R. Guarini, F. Battisti, and A. Chiovitti, "A methodology for the selection of multi-criteria decision analysis methods in real estate and land management processes," *Sustainability (Switzerland)*, vol. 10, no. 2, 2018.
- [18] C. Bana e Costa (2018). Multi-criteria Value Measurement, lecture notes. Retrieved from: https://fenix.tecnico.ulisboa.pt/downloadFile/1689468335616940/MAD%202018_2019%20Larry.pdf (Accessed: 06-05-2019).
- [19] C. Bana e Costa, J.-M. De Corte, and J.-C. Vansnick, "MACBETH. (Overview of MACBETH multicriteria decision analysis approach)," *International Journal of Information Technology and Decision Making*, vol. 11, no. January, pp. 359–387, 2003.
- [20] C. A. Bana e Costa, J. A. Ferreira, and E. C. Correa, "Metodologia multicritério de apoio à avaliação de propostas em concursos públicos," 2000.
- [21] T. Cardoso, M. D. Oliveira, A. Barbosa-Póvoa, and S. Nickel, "Moving towards an equitable long-term care network: A multi-objective and multi-period planning approach," *Omega (United Kingdom)*, vol. 58, pp. 69–85, 2016.
- [22] C. A. Bana E Costa, M. C. Carnero, and M. D. Oliveira, "A multi-criteria model for auditing a Predictive Maintenance Programme," *European Journal of Operational Research*, vol. 217, no. 2, pp. 381–393, 2012.
- [23] T. C. Rodrigues, "The MACBETH Approach to Health Value Measurement: Building a Population Health Index in Group Processes," *Procedia Technology*, vol. 16, pp. 1361–1366, 2014.
- [24] P. R. Pinheiro, A. K. A. De Castro, and M. C. D. Pinheiro, "A multicriteria model applied in the diagnosis of Alzheimer's disease: A Bayesian network," *Proceedings - 2008 IEEE 11th International Conference on Computational Science and Engineering, CSE 2008*, pp. 15–22, 2008.
- [25] D. F. Lopes, M. D. Oliveira, and C. A. Bana E Costa, "Occupational health and safety: Designing and building with MACBETH a value risk-matrix for evaluating health and safety risks," *Journal of Physics: Conference Series*, vol. 616, no. 1, 2015.
- [26] A. Miquel-Cases, P. C. Schouten, L. M. Steuten, V. P. Retèl, S. C. Linn, and W. H. van Harten, "(Very) Early technology assessment and translation of predictive biomarkers in breast cancer," *Cancer Treatment Reviews*, vol. 52, pp. 117–127, 2017.
- [27] K. Marsh, E. Zaiser, P. Orfanos, S. Salverda, T. Wilcox, S. Sun, and S. Dixit, "Evaluation of COPD Treatments: A Multicriteria Decision Analysis of Aclidinium and Tiotropium in the United States," *Value in Health*, vol. 20, no. 1, pp. 132–140, 2017.
- [28] A. C. Vieira, M. D. Oliveira, and C. A. Bana e Costa, "Enhancing knowledge construction processes within multicriteria decision analysis: The Collaborative Value Modelling framework," *Omega (United Kingdom)*, pp. 1–15, 2019.
- [29] N. Slocum, *Participatory methods toolkit: A practitioner's manual*. 2003.
- [30] T. C. Rodrigues, M. D. Oliveira, C. A. Bana e Costa, and P. Santana, "The MACBETH approach to health value measurement : a multicriteria model for building a value-based population health index," 2014.
- [31] "WP6 Decision Support for Multicriteria Modelling of the Population Health Index and Evaluation, Foresight and Selection of Policies". Euro-Healthy. Available at: <http://www.euro-healthy.eu/research/decision-multicriteria-modelling> (Accessed: 16-09-2019).