TÉCNICO LISBOA

Hospital-based health technology assessment of emerging medical devices:

A socio-technical approach for multicriteria evaluation of emerging biomarkers for HER2+ breast cancer

Beatriz Leitão Gomes Coelho

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Professor Mónica Duarte Correia de Oliveira
Doctor Hugo Miguel Lobato Quintino

Examination Committee

Chairperson: Professor Mário Jorge Costa Gaspar da Silva
Supervisor: Professor Mónica Duarte Correia de Oliveira
Member of the Committee: Professor Teresa Sofia Cipriano Gonçalves Rodrigues

July 2021

Preface

The work presented in this thesis was performed at Centro de Estudos de Gestão of Instituto Superior Técnico (IST), Universidade de Lisboa (Lisbon, Portugal), during the period of February 2020 to July 2021, as part of the MEDI-VALUE project and under the supervision of Professor Mónica Duarte Correia de Oliveira, from Instituto Superior Técnico (IST), and co-supervised by Doctor Hugo Miguel Lobato Quintino, from Hospital do Espírito Santo de Évora (HESE).

Declaration

I declare that this document is an original work of my authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I would like to thank my supervisors: Professor Mónica Oliveira, not only for providing guidance and feedback throughout this master thesis but also for always having confidence in this work; and Doctor Hugo Quintino, for the availability and motivation to participate in this project.

I would also like to thank all the healthcare professionals at Hospital Espírito Santo de Évora that were willing to participate in various steps of my thesis, in particular Doctor Jorge Caravana, for his insight and expertise.

A special thank you goes to Liliana Freitas, for her work as a member of the facilitation team for this thesis, and whose support as part of her Ph.D. was indispensable in the development of this project. Thank you for being there when I needed it the most and for all the meetings and conversations!

I am also thankful to Professor Carlos Bana e Costa, for his insights and constructive criticism.

Last, but not least, a big thank you goes to my friends and family, especially my mother, father, and brother, for their unconditional love, support, and for always being my biggest cheerleaders. To João, for his positive energy: thanks for keeping me sane!

Abstract

Health systems are facing significant challenges, with new medical devices and technologies constantly entering the market. Consequently, there is an increasing need to assess their quality, safety, effectiveness and inform decision-makers on the adoption of these technologies. In this context, Health Technology Assessment (HTA), combined with innovative Multicriteria Decision Analysis (MCDA) models and tools, have the potential to face some of these challenges.

Biomarkers are a good example of an innovative medical device. However, despite the current progress with the discovery of multiple new biomarkers, many of them are still not used in clinical practice and are yet to be approved by international guidelines.

Thus, and within the scope of Hospital-based HTA, this thesis has the core objective of developing and implementing an innovative socio-technical approach, combining the MACBETH approach for multicriteria evaluation with a rapid review of evidence and with specially designed participatory processes, using the WisedOn software as a decision support system. The approach will enable the multicriteria evaluation of emerging biomarkers, for HER2+ breast cancer, to assist decision-making at Hospital do Espírito Santo de Évora (HESE).

With this methodology it was possible to obtain a comprehensive evaluation of the emerging biomarkers. Despite the initial challenge of lack of information, the results show the approach was useful to achieve a consensus among the several key-actors, which resulted in a model with six criteria for the evaluation of eleven emerging biomarkers. The model also has the potential to be reused in other contexts in HESE, for other health technologies.

Keywords: Hospital-based Health Technology Assessment • Multicriteria Decision Analysis • Socio-technical approach • MACBETH • Emerging biomarkers • HER2+ breast cancer

Resumo

Atualmente os sistemas de saúde enfrentam múltiplos desafios, associados com a entrada de novas tecnologias e dispositivos médicos no mercado. Consequentemente, existe uma necessidade crescente de avaliar a sua qualidade, segurança, eficácia e de informar no processo de tomada de decisão sobre a adoção destas tecnologias. Neste contexto, a Avaliação de Tecnologia em Saúde, aliada a modelos e ferramentas inovadoras de Análise Multicritério de Apoio à Decisão, têm o potencial para enfrentar alguns destes desafios.

Os biomarcadores são um bom exemplo de dispositivos médicos inovadores. Contudo, apesar do atual progresso, muitos deles ainda não são usados na prática clínica e não se encontram aprovados por diretrizes internacionais.

Assim, esta dissertação tem como objetivo principal desenvolver e implementar uma abordagem sóciotécnica inovadora, combinando a abordagem MACBETH para avaliação multicritério com uma revisão rápida da evidência e com processos participativos, aliados ao software WisedOn como sistema de suporte à decisão. A abordagem permitirá a avaliação multicritério de biomarcadores emergentes para o cancro da mama HER2+, auxiliando assim a tomada de decisão no Hospital do Espírito Santo de Évora (HESE).

Seguindo esta metodologia, foi possível obter uma avaliação abrangente dos biomarcadores emergentes. Apesar do desafio inicial relativo à falta de informação, os resultados mostram que a abordagem foi útil para alcançar um consenso entre os diversos atores-chave, o que resultou num modelo com seis critérios para a avaliação de onze biomarcadores emergentes. O modelo apresenta também potencial para ser replicado noutros contextos, para avaliar outras tecnologias no HESE.

Keywords: Avaliação de Tecnologias da Saúde em ambiente hospitalar • Modelo Multicritério de Apoio à Decisão • Abordagem sociotécnica • MACBETH • Biomarcadores emergentes • Cancro da mama HER2+

Contents

List of Tables

List of Figures

Acronyms

AUC Area Under the Curve. 19, 20

CA Competent Authority. 5

CE Conformité Européenne. 5

ER Estrogen Receptor. 23, 25, 26

EU European Union. xi, 3–6, 15

FcγRs Fragment crystallizable-gamma receptors. 29, 42, 55

FDA Food and Drug Administration. 10

FPV Fundamental Point of View. 35, 36

HB-HTA Hospital-Based Health Technology Assessment. 2, 13–18, 69, 73, 74

HER2 Human Epidermal Growth Factor Receptor 2. 11, 23, 25, 27–29, 40–42, 53, 54, 58, 66

HER2+ breast cancer Human epidermal growth factor receptor 2 positive breast cancer. xi, xv, 2, 13, 27–29, 31–33, 40–47, 53–57, 66, 69–71, 73

HER3 Human Epidermal Growth Factor Receptor 3. 41, 42, 53, 54

HESE Hospital do Espírito Santo de Évora. xi, xv, 2, 13, 32, 33, 40–43, 46, 48, 51, 53, 54, 56, 58–60, 66, 67, 69, 70, 73, 74

HR Hormone Receptor. 29

HTA Health Technology Assessment. 1, 2, 11, 13–19, 21, 31, 69, 70, 73, 74

IVD *In Vitro* Diagnostic. 6, 7, 11

MACBETH Measuring Attractiveness by a Categorical Based Evaluation Technique. xv, 2, 21, 29, 31, 32, 34, 37–39, 47–49, 51, 63, 69, 70

MCDA Multiple Criteria Decision Analysis. 13, 17, 20, 21, 31, 34, 48, 69, 70, 73, 74

MEDI-VALUE Developing HTA tools to consensualise MEDIcal devices' VALUE through multicriteria decision analysis. 1, 2, 32, 42, 43, 46, 56, 69, 73, 74

NB Notified Body. xi, 5–7

OECD Organization for Economic Cooperation and Development. 1

PAM50 Prediction analysis of microarray 50. 54

PD-L1 Programmed death-ligand 1. 29, 42, 55, 58, 65, 66, 69

PI3K Phosphoinositide 3-kinase. xvi, 42, 54, 58, 65, 66, 69

PIK3CA Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha. xvi, 29, 42, 54, 58, 66, 69

PR Progesterone Receptor. 23, 25, 26

PTEN Phosphatase and tensin homolog. 42, 54, 66, 68

PV Point of View. 34, 35

ROC Receiver Operating Characteristic. 19, 20

TIL Tumour-Infiltrating Lymphocyte. xvi, 29, 42, 55, 65, 66, 69

USA United States of America. 3, 4, 10

WHO World Health Organization. 13

Chapter 1

Introduction

1.1 Motivation

Technology has globally changed health care systems and has increasingly become a dominant force. New medical and clinical technologies are currently being developed, introduced in international markets (and consequently health care systems) at a rate never seen before [1]. Moreover, the development of personalized drugs used in precision medicine therapeutics, robotics, 3D printing, nanotechnology, artificial intelligence, or stem cells is steadily, and already, shaping the future of health care [1].

Despite the apparent positive revolution happening in front of our eyes which is providing recent and state-of-the-art drugs, medical devices, or diagnostic procedures, new challenges also arise. A common challenge is associated with the need for regulation reforms since oftentimes recent technologies combine features of multiple technologies (*e.g.* a medical device with a drug delivery system) that are evaluated by separate entities or due to the necessity of performing new clinical trial models for drugs developed for precision medicine [2].

The quality, safety, and efficacy of new technologies must always be assessed before they enter the market and is key to inform health decision-makers on the adoption of new technologies, in the clinical context. Besides that, constant monitoring of these technologies must also be conducted, after market entry, so that only technologies that provide value to the health care systems are used [3].

To assure patient safety and monitoring of new technologies, countries possess entities and agencies responsible for the assessment and evaluation of health care technologies (*e.g.* medical devices). Nevertheless, despite the apparent strict regulations and health technology assessments performed by these entities, in order to assess technologies in terms of their benefits, risks, and costs, evidence points out that the majority of medical devices enters the health care system are adopted and used by clinicians with little or sometimes no health gain associated [2].

In 2017, the Organization for Economic Cooperation and Development (OECD) published a report regarding new health technologies [2] and stated that many new biomedical technologies are approved in the market with small evidence of their effectiveness and safety, claiming that the assessment of these devices, under real-world conditions, is scarce [2]. Other studies similarly conclude that there is a lack of transparency regarding clinical evidence for medical devices worldwide, and consequently, having healthcare professionals using medical devices without real knowledge or data regarding them [3].

With that being said, this thesis intends to develop methods and tools to provide healthcare professionals with synthesized medical device evidence of its benefits, risks, and costs and to explore health professionals' interpretation and need of further evidence and information. This thesis is inserted in the Developing HTA tools to consensualise MEDIcal devices' VALUE through multicriteria decision analysis (MEDI-VALUE) project [4], based on Health Technology Assessment (HTA) to assess

the overall value of medical devices, valuable information that can improve the current health systems, promoting both innovation and sustainability, not only for clinical care and overall population health but also for research and development. The MEDI-VALUE project combines multiple partners: in the academic field, the Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento (IST-ID) and the London School of Economics and Political Science (LSE). Autoridade Nacional do Medicamento e Produtos de Saúde, I.P. (INFARMED), the Portuguese national HTA agency and the hospitals of Centro Hospitalar Lisboa Norte, HESE and Instituto Português de Oncologia de Lisboa (IPO Lisboa) are also part of this project, providing the health needs [4].

Thus, and within the scope of Hospital-Based Health Technology Assessment (HB-HTA), the main goal of this project is to develop and implement a socio-technical approach to construct a multicriteria model. This approach combines the use of the MACBETH approach with a rapid review and participatory methods (results from a prior Web-Delphi process, interviews, Workshops and Decision Conferences) allied to the WisedOn decision support system to create a multicriteria evaluation of biomarkers for breast cancer, more specifically, for the HER2+ breast cancer subtype.

Consequently, it is expected that this thesis can fulfill the need for evidence synthesis and also provide relevant information and data to the HESE healthcare professionals, in order to aid the decision-making process. Additionally, the model can be reused since it can be used by HESE to assess other emerging biomarkers, not only for breast cancer In fact, emerging biomarkers for other diseases can be assessed to be used in future clinical practise.

1.2 Document organization

Besides the introductory chapter, this thesis is organized in other six distinct chapters, namely:

- Chapter 2, which provides some background and concepts associated with medical devices, with a focus on biomarkers, fundamental for the understanding of this thesis;

- Chapter 3, concerning a deep review regarding HTA (including HTA of biomarkers) and breast cancer, including breast cancer epidemiology, classification, biomarkers' use, and the specific case of HER2+ breast cancer disease;

- Chapter 4, describes in detail the proposed methodology, based on a socio-technical approach, developed in this thesis;

- Chapter 5, presents the results obtained using the proposed methodology;

- Chapter 6 presents a brief discussion, not only regarding the results but also this thesis as a whole;

- Chapter 7, lastly, that includes the conclusion, as well as some reflections for future work.

Chapter 2

Context

In this chapter, a brief overview of some fundamental concepts is proposed, for a better understanding of this thesis.

First, an overall description of medical devices will be presented, including the markets in which they are present and their characteristics, medical device classification, and the current regulations applied; a focus on the *in vitro* diagnostic medical devices, in which biomarkers are inserted, will also be done. This will allow to then introduce and cover the topic of biomarkers, particularly, its life cycle, classification, and their use in personalized medicine.

2.1 Medical Devices

According to the European Parliament and the Council of the European Communities, a medical device consists of: "*any instrument, apparatus, appliance, material or other article, whether used alone or in combination, including the software necessary for its proper application intended by the manufacturer to be used for human beings for the purpose of:*

- *diagnosis, prevention, monitoring, treatment or alleviation of disease;*
- *diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap;*
- *investigation, replacement or modification of the anatomy or of a physiological process;*
- *control of conception*

and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means (...)" [5] This extensive definition translates the variety and diversity of medical devices, that can range from the most common and simple devices, such as disposable syringes or gloves, to implants or *in vitro* diagnostic tests, and the most complex ones being imaging equipment and robotics or life-sustaining ones, such as a cardiac defibrillator.

Each country (or group of countries) has its own medical device rules and regulations concerning the marketing of medical devices. For example, the regulations in the United States of America (USA) differ from the ones in China, Mexico, or Canada. Thus, it is important, for a medical device company, not only to study the market but also the country in which is entering, in order to comply with the current country's regulations.

For the purpose of this thesis, a focus will be performed solely on the case of medical devices in the context of the EU.

2.1.1 Market

The total number of medical devices available worldwide is growing and is expected to continue to grow. This translates, of course, into the medical device market's dimensions. The global medical device market size is constantly rising and, what was in 2018 a USD 425.5 billion market (approx. €393 billion), is expected to rise to USD 612.7 billion (approx. €567 billion) by 2025 [6].

According to the medical device type, the market dynamics will vary. For manufacturers mainly focused on the most simple medical devices, *i.e.* routine surgical devices, the market competition is high (due to the low risk associated with the products), mainly relying on high sales volumes of the product to become profitable [7]. However, markets for high-risk medical devices present lower competition, are more difficult to enter, and usually perform opaque pricing strategies (*i.e.* products are sold at a hidden, low price, by the companies). So, unlike the previous market, in this one, companies can charge higher prices and consequently have higher profit margins, usually ranging from 20% to 30% [7].

Considering the EU case, it was estimated that the medical technology market was, at roughly, €115 billion, in the year 2017 and that, based on manufacturer prices, the EU medical technology market is estimated to make up around 27% of the world market, the second-largest market, after the USA (with around 43% of the global market) [8].

Another interesting fact relies on the expected areas of growth. A study developed in 2018 predicted that, on a worldwide perspective, it is expected that *in vitro* medical devices will continue to be the largest medical device sector, with annual sales around USD 79.6 billion (around €66.92 billion) and a 13.4% market share of the medical device industry (Figure 2.1); followed by cardiology and diagnostic imaging (despite a slower growth regarding this last medical device segment) [9].
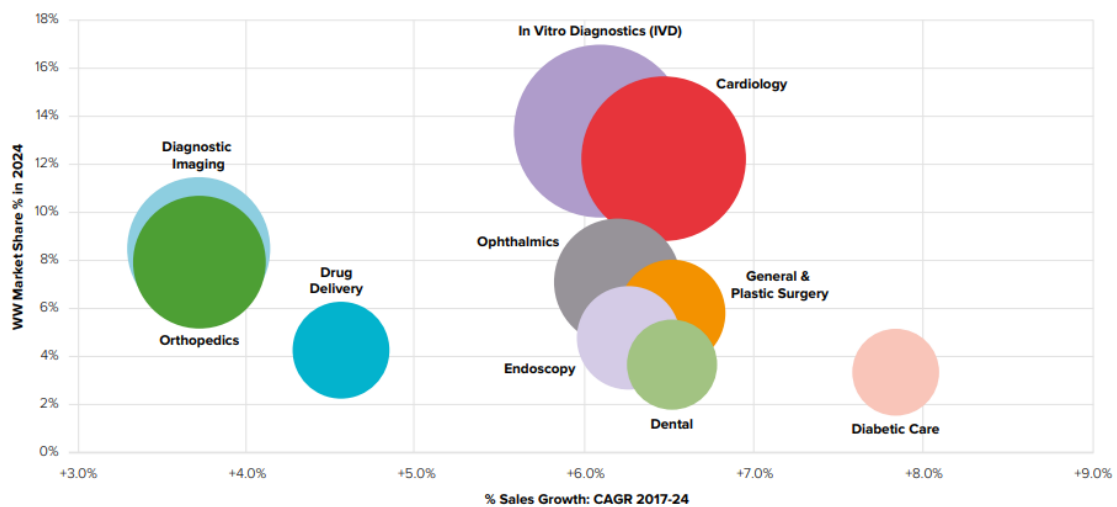
Figure 2.1: Different areas of the world medical technology market, by market share and sales growth (2017-2024). [source: [9]]

2.1.2 Medical Device classification

In the EU, according to the classification rules listed in Directive 93/42/EEC Annex IX [10], all medical devices are placed into one of four categories. Since medical devices are associated with different risks (upon its use), it was considered more feasible, economical, and justifiable to categorize them, rather than to place them all under rigorous conformity assessment procedures.

Therefore, medical devices are divided into three distinct classes, with increasing risk: class I, II, and III. Furthermore, class II devices are also divided into IIa and IIb categories (*i.e.*, class I, IIa, IIb,

or III, with increasing user risk - Table 2.1) [10]. This classification is given by the manufacturer, before entering the market.

Table 2.1: Risk categories and evidentiary requirements for medical devices in EU. [adapted from: [2]]

Risk stratification	Evidence required
Class I: No to negligible risk (*e.g.* thermometers)	No approval (self-certification) with clinical evaluation required.
Class IIa: Low risk (*e.g.* infusion pumps)	Dossier of supporting literature to substantiate safety and performance comprising clinical and non-clinical data. Assessment by notified body, which involves an audit of the Quality Management System of the manufacturer's production processes.
Class IIb: Medium risk (*e.g.* dialysis machines, artificial joint)	
Class III: High risk (*e.g.* pacemakers)	Clinical studies required, can be non-randomised and single arm, focused on demonstrating safety. Assessment of the study design and of clinical evidence by a notified body is required.

2.1.3 Current Regulation and the new Medical Devices Regulation (MDR)

For any health technology, such as drugs or medical devices, a regulatory system is required, *i.e.*, a set of rules that prevent or, at least, limit the risk of a product producing harm and not fulfilling its overall objective [11]. By doing so, it is possible to provide quality, safe and effective technology to the patients. Generally, a national governing body is responsible to create the necessary rules, apply them in the nation, and assure that the law is enforced. The laws reach all the manufacturers, sellers, and final users of the technology, with the latter being in the majority of cases, health care professionals (such as surgeons or clinicians). Unlike the drug regulatory system, the medical device one is far less stringent [7]. However, it is still necessary to impose a more rigorous regulation for this type of technology, not only due to its higher variability in terms of number and risk types but also because of the constant incremental improvements that are added to already launched medical devices, which usually do not happen so often in the drug's markets.

As already mentioned, different countries have distinct regulations and classification rules, that are not always uniform and comparable, due to their variation from country to country. Given the complexity of the system, only the EU case will be explained in further detail.

Thus, in the EU, the regulations are currently based on three distinct directives (initially introduced in the 1990s), that regulate the safety and marketing of medical devices [10] [12]:

- Directive 93/42/EEC - Medical Devices Directive (MDD);
- Directive 90/385/EEC - Active Implantable Medical Devices (AIMD);
- Directive 98/79/EEC - *In Vitro* Diagnostic Medical Device Directive (IVDD).

In agreement with these regulations, for a medical device to enter European markets, it must present a Conformité Européenne (CE) mark. A CE mark corresponds to a manufacturer's declaration that translates the compliance of the product, according to the essential requirements [2].

In this process, two authorized representatives are involved: Competent Authority (CA) and NB [10]. Simply put, CAs, that exist in each European member state, designate NBs (external organizations), that conduct assessments to verify if the medical device is under the EU Directive requirements. This is done by conducting audits of the applier's quality system and verifying the evidence provided by the manufacturer's technical reports, in regards of its safety and effectiveness. Once the medical device is approved for a CE mark, it can be freely marketed in the European Economic Area (EEA) without further control [10].

Due to new and emergent medical technologies, that exposed certain fragilities and pointed to limitations in the designed framework, these three directives had to be complemented by several updates throughout the years [10].

Thus, to homogenize EU market access for medical devices, to face the current market entry of medical devices associated with low clinical evidence, and to improve post-market surveillance, a new legislation, published on May 5th, 2017 in the Official Journal of the European Union [5], were introduced, replacing the three current and decades-old medical device directives. These new requirements were adopted by the medical device manufacturers until 26th May 2020 and gave rise to two new distinct Regulations:

- Regulation 2017/745 on Medical Devices (MDR);
- Regulation 2017/746 on In-Vitro Diagnostic Medical Devices (IVDR).

However, during the transitional period, NBs could still certify and introduce into the European market, according to the previous directives. The Medical Device Regulation (MDR) was in full application by May 2020, while the *In Vitro* Diagnostic Regulation (IVDR) will be two years later, in May 2022 [13].

With this new regulation, it is expected that some developments and improvements will arise, when compared with the current medical device regulation. First, the regulation will be more homogeneous for the medical devices available in the EU market, with a considerable product scope expansion (*e.g.* devices without a direct medical purpose, used for aesthetics purposes, such as cosmetic implant devices or coloured lenses will now also be considered under this regulations) [13].

Besides that, these new regulations will also be stricter and will demand rigorous clinical evidence (and will be, therefore, more similar to the pharmaceutical products' regulations). This is a fundamental and required characteristic for health technologies, in order to assure quality, safety, and risk-free medical devices for its patients and users.

To do so, under these new regulations, each medical device will be associated with a unique identifier (unique device identifier - UDI) [7], that will facilitate the ability to identify and track down the origin of the device (manufacturer, economic responsibilities, materials, and overall traceability). In order to ease the access and to identify medical devices, a European database on medical devices (Eudamed) [5] will also be created, so that any healthcare professional and the public can have total and transparent access to it. It will also be possible, with this new system, to know the exact number of medical devices available in the market, a number that, up to date, is still uncertain.

Due to the Covid-19 pandemic, that unexpectedly took the world by surprise, the date for the new Medical Device Regulation (MDR) was postponed one year, to 26th May 2021. This will remove pressure from national authorities, notified bodies, manufacturers, and other actors involved in the process, so that they can focus on urgent priorities associated with the coronavirus crisis. Nevertheless, the new *In Vitro* Diagnostic Regulations (IVDR) will maintain its date of application to 26th May 2022, until further update [5] [14].

2.1.4 The specific case of *In Vitro* Diagnostic Medical Devices

In Vitro Diagnostic (IVD) medical devices consist on a specific subtype of medical devices and has, therefore, specific characteristics and regulations associated. Based on the European Parliament and the Council of the European Union, an *in vitro* diagnostic medical device consists of *"any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, piece of equipment, software or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of specimens, including blood and tissue donations, derived from*

the human body, solely or principally for the purpose of providing information on one or more of the following:

- *concerning a physiological or pathological process or state;*
- *concerning congenital physical or mental impairments;*
- *concerning the predisposition to a medical condition or a disease;*
- *to determine the safety and compatibility with potential recipients;*
- *to predict treatment response or reactions;*
- *to define or monitoring therapeutic measures." [5]*

Just like other medical devices, IVD medical devices must also comply with regulatory requirements.

IVD medical devices are a large and varied sector of medical devices, with a distinct impact on the diagnostic and treatment of patients. Just like regular medical devices, IVDs are also classified according to the risk they impose: an IVD with higher risk associated must also have a more stringent assessment and vice-versa. However, for the specific case of IVDs, its risk is indirect and associated with the risk of an incorrect diagnosis (*i.e.,* risk of an erroneous result from the assay) not only for the patient but for the population in general [15].

That is why the rules for the classification of general medical devices, based on interactions with the body, are not suitable for this subclass of medical devices. Under the current directive, IVDs are laid down in two restrictive lists: List A and List B. IVDs in List A are the highest risk devices, and therefore, demand a more stringent and extensive examination by the NB when compared with the IVDs on List B [16]. Some IVDs are not listed in any of these lists and can be referred to as Devices for Self-Testing and Self-Declared Devices, both not requiring assessment by a notified body [8].

Due to the use of restrictive lists, with a specific number of IVDs, it implies that newly developed technology that is not mentioned in this list is, by default, not obliged to go under the scrutiny of a NB [16]. This is an issue of the current directive and it will be resolved with the new regulation for *in vitro* diagnostic medical devices (IVDR).

Under this new regulation, a new risk-based classification system will be used that is associated with the risk that the device presents to the population in general (in the case of a serious infectious disease), or the individual (diagnostic of cancer), as a result of an incorrect test result. With this system, IVDs will be classified from class A to D, with class D being the highest risk class (Table 2.2) [16].

Table 2.2: Risk categories based on the *in vitro* diagnostic medical devices Regulation (IVDR), with respective NB involvement and examples. Increasing risk and increasing NB involvement, from A to D. [adapted from: [16]]

Certification	IVDR class	Examples
Notified body approval required	D	• Hepatitis B blood-donor screening • HIV test • ABO blood grouping
	C	• Blood-glucose self-testing • Cancer markers • Human genetic testing
	B	• Pregnancy self-testing • Urine-test strips • Cholesterol self-testing
Self-certification	A	• Specimen receptacles • Clinical chemistry analysers

Currently, and as already referred, the *in vitro* diagnostic sector is regulated by Directive 98/79/EEC (the *in vitro* diagnostic medical device Directive). However, from 26th May 2022 onwards, the Regulation 2017/746 on *In Vitro* Diagnostic Medical Devices (IVDR) will fully apply. During this transitional period, manufacturers are free to comply with either the Directive or the Regulation [8].

2.2 Biomarkers

Based on the Biomarkers Definition Working Group, a biological marker, or biomarker, consists of "*a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*" [17]. Biomarkers can also be measured from a variety of origins: tissue, stool, or body fluids, such as blood (or plasma) and urine.

Biomarkers have been increasingly studied throughout the years and the progress made in a variety of fields such as genomics, genetics, proteomics, and imaging has ultimately highlighted the importance of biomarkers as useful clinical indicators in various diseases, including breast cancer [18]. Biomarker literature has exploded in recent years [19]. In fact, a simple search on the PubMed database [20] can confirm this growth: in this database and 2019 alone, more than 68 000 reports were published, comparing with less than 15 000 in the year 2000. This "exponential" growth is clear in Figure 2.2.
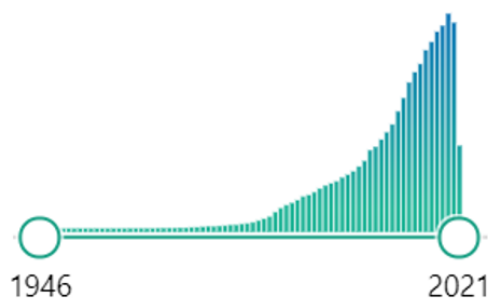
Figure 2.2: Search results in PubMed database for the term "biomarker", ranging from 1946 to May 2021. [source: [20]]

2.2.1 Life cycle of a biomarker

Biomarkers do not arrive in preclinical and clinical practice fully developed and validated, with applicability for all the population and with ideal guidance for its use [18]. In fact, for their discovery and consequent development, a large investment regarding both time and resources must be performed [19]. Bellow, the biomarker discovery and development pipeline is represented (Figure 2.3).
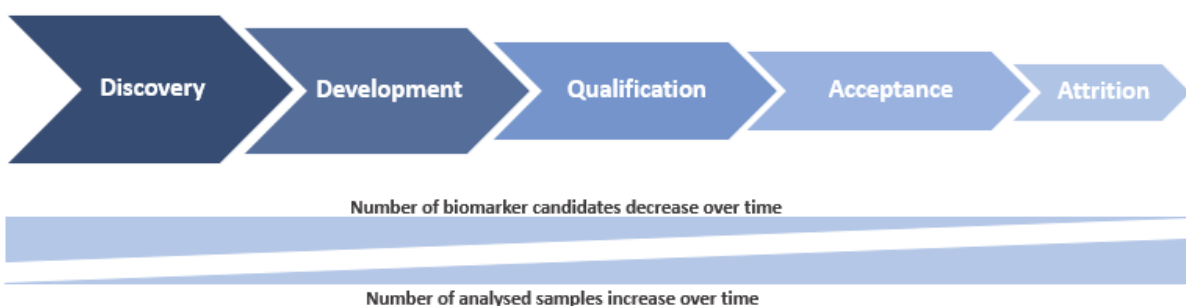
Figure 2.3: Biomarker discovery and development pipeline. [adapted from: [19]]

The distinct phases associated with the biomarker pipeline are:

1. **Discovery** - Consists of the first phase of the pipeline. In this phase, biomarker candidates are discovered, both in clinical and non-clinical studies, that include DNA variations, transcriptome, proteome, phosphoproteome, metabolome, circulating cells history features, and imaging properties, amongst others. Typically this stage lasts between 1 to 3 years [19].

2. **Development** - The development phase, which lasts between 3 to 5 years, is divided into 2 main parts [19]:

   - Analytical Validation, during which the results must show not only reproducibility across the different samples but also reproducibility and standardization of the assay;
   - Clinical or Non-Clinical Utility Studies, in which the results must show performance characteristics, well-designed experiments and bring added value in research models and/or to the patients.

3. **Qualification** - In the third stage of the pipeline is when the biomarker is tried for qualification, *i.e.*, it requires agreement about the intended context of use by the multiple stakeholders; coordination of efforts towards data generation, and strong evidence for the stated context of use. If the biomarker is able to go to the next phase (*i.e.*, is qualified), it will be accepted in future drug programs, without re-review. This phase typically lasts less than 1 year [19].

4. **Acceptance** - This phase is related to acceptance in a drug program, that typically depends on the quality of evidence showing utility; reproducibility in well-designed studies; technical abilities of outside groups, and the dissemination of data [19].

5. **Attrition** - The last phase is the attrition phase that may occur if the biomarker's performance was misjudged due to new evidence that appeared related with limited efficacy of the biomarker or data showing possible harm. This is a phase to avoid since it is usually associated with waste of time and money [19].

As already referred, the biomarker discovery and development pipeline is a very long and expensive process.

While the excitement and innovation continue to move the biomarker science forward, it is still important to keep in mind that many biomarker studies report promising and relevant results, without proceeding to confirm these findings in more elaborate study types (*e.g.* independent multi-institutional cohorts). Adding to this, it has also been reported that many biomarkers studies may be associated with inappropriately validated reagents and assays, poorly collected and stored samples, inadequate study design, publication bias, etc [19]. Lastly, from a large number of biomarkers in the beginning (the emerging biomarkers) and despite their high potential for clinical application, only a fraction of them will be reaching the end of the narrow bottleneck (Figure 2.3), leading to high losses of resources not only for investors but also for the general public.

Thus, a collaborative effort must be performed, involving multiple stakeholders, including the industry (companies), regulators, and disease foundations not only to catalyze faster approvals, with safe and effective biomarkers but also to reduce the costs associated.

2.2.2 Biomarker classification and use

There is great variety when referring to biomarkers. As already mentioned, they can be measured from different origins from the body, using a variety of procedures (physical, optical, or enzyme assay; immunochemistry; mass spectrometry; cytological and histochemical analyses; or by *in vivo* imaging

techniques) [21]. This diversity also extends to the molecular type of the biomarker. Therefore, the biomarker can be related with [21]:

- DNA on the genomic level (copy number variations, gene amplification, gene mutations, genetic polymorphisms, and disease-related methylation patterns);

- RNA transcripts (expression levels and alternative splicing products, expression of noncoding transcripts as RNAi or microRNA molecules);

- Proteins (expression levels, altered activities, altered localization, post-translational modifications);

- Lipids and other metabolites (molecular identity and concentration).

Biomarkers can be used as precise measurement tools to determine disease progression and the effects of a specific intervention (such as surgery, drugs, or vaccines) [17].

Due to their vast diversity and application in different contexts, there are many ways to classify biomarkers, leading to a lack of clarity and ambiguous use of the terms. Thus, in order to harmonize the terminology used in the biomarker field, the Food and Drug Administration (FDA), a federal agency in the USA responsible for the regulation of medical devices, drugs and other products, classified the distinct biomarkers used in clinical context in 7 categories [22] [23]:

- **Diagnostic biomarker** - used to detect or confirm presence of a disease or condition of interest or to identify individuals with a subtype of the disease;

- **Monitoring biomarker** - measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent;

- **Pharmacodynamic/Response biomarker** - used to show that a biological response has occurred in an individual who has been exposed to a medical product or an environmental agent;

- **Predictive biomarker** - used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent;

- **Prognostic biomarker** - used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest;

- **Safety biomarker** - measured before or after an exposure to a medical product or an environmental agent to indicate the likelihood, presence, or extent of toxicity as an adverse effect;

- **Susceptibility/Risk biomarker** - indicates the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or the medical condition.

2.2.3 Biomarkers in personalized medicine

Personalized medicine, or precision medicine, has rapidly grown throughout the years, shifting the traditional clinical practice into a new paradigm [24] [25]. It is an emerging field of medicine that consists of tailoring the medical treatments given to a patient based on his individual characteristics. Therefore, this approach relies on scientific breakthroughs in the understanding of how an individual's unique genetic and molecular profile makes him susceptible to a specific disease [26].

A new paradigm was created, that transformed the traditional "one-size-fits-all" approach, in which patients are treated with the same drugs, designed to the general population in a more individualized approach. It has clear patient benefits that one cannot assure in a traditional approach: not only does it improve the opportunity of disease prevention and treatment options, but it also helps to avoid adverse

drug reactions. Besides, it can also increase treatment options, improve administration methods and select an optimal therapy, for overall improved quality of life [25].

Personalized Medicine is a multifaceted approach to patient care, encompassing [25]:

- Risk Assessment - genetic assessment to reveal a predisposition to a disease;
- Prevention - behaviour/lifestyle/treatment intervention to prevent a disease;
- Detection - early detection at a molecular level;
- Diagnosis - accurate disease diagnosis enabling individualized treatment strategy;
- Treatment - improved outcomes through targeted treatments and reduced side effects;
- Management - active monitoring of treatment response and disease progression.

Beyond the patient benefits already mentioned, personalized medicine also brings benefits to the healthcare system itself, namely with the potential to increase patient compliance, shift the overall goal of medicine from reaction to prevention, improve the current cost-effectiveness associated with drugs and therapeutics used and increase patient confidence in the process [25] [26].

Biomarkers in personalized medicine are extremely relevant. In fact, they are intimately related: due to the use of biomarkers it is possible to improve molecular diagnostics and to consequently associate a specific therapy, according to the diagnostic. Besides, with biomarkers, it is also possible to monitor the response to therapy. Last, but not least, biomarker diagnostics help to target the right drugs for the right patient, including the discovery and development of these drugs (Figure 2.4) [25] [27].
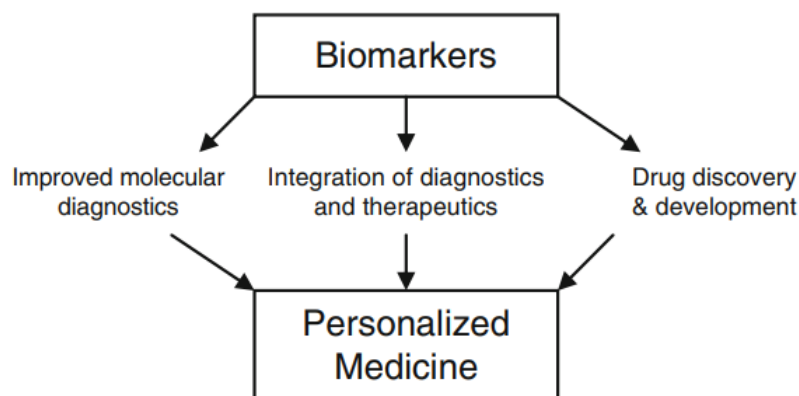
Figure 2.4: Role of biomarkers in personalized medicine. [source: [27]]

Accompanying the rapid growth of personalized medicine and the number of biomarkers available, IVD applicable biomarkers tests are also growing [28]. The success of personalized medicine ultimately resides on the development of biomarkers and diagnostic tests (*in vitro* diagnostic medical devices) that can be used to accurately identify the patient population that will benefit from specific treatments.

Regarding the specific case of breast cancer disease, and since it is a very heterogeneous disease, making it difficult to classify the patients and to introduce personalized therapeutic approaches due to the various disease subtypes, new and emerging biomarkers associated with one of this diseases subtypes will be assessed in this thesis. Thus, in the following chapter, a literature review is presented, concerning HTA and breast cancer, including the current biomarkers associated with this disease as well as the breast cancer subtype that will be the focus of this work, the Human Epidermal Growth Factor Receptor 2 (HER2) subtype.

Chapter 3

Literature Review

In this chapter, a literature review will be presented, focusing on HTA, the role of Multiple Criteria Decision Analysis (MCDA) in HTA, and breast cancer, with an emphasis on the HER2+ breast cancer subtype.

First, a review on HTA will be performed, including HB-HTA, the collaboration between HTA agencies and HB-HTA, the currently existing gap between clinical research and clinical practice, and some examples of HTA carried out for biomarkers, including effectiveness and economic evaluations. Besides, a brief explanation of the relation between HTA and MCDA is presented.

With the support of HESE, it was clear from an early stage during the initial meetings that there was the need to explore which emerging biomarkers could be adopted and used in clinical practise so that a better management of breast cancer patients could be done. For that, in this chapter, the breast cancer disease will be described, encompassing not only the disease epidemiology but also its classification on the different subtypes, available treatments, and the current biomarkers used in this disease. Lastly, a deeper explanation of the HER2+ breast cancer subtype will be presented, due to its importance in this thesis.

3.1   Health Technology Assessment (HTA)

To understand which technologies have value and are relevant to the current health systems, both preserving and increasing the quality of the existing health systems, while reducing the increasing health costs associated with it, a variety of jurisdictions, driven by policy makers and clinicians, is becoming more interested in HTA.

According to the World Health Organization (WHO), HTA is *"the systematic evaluation of properties, effects and/or impacts of health technologies and interventions. (. . .) The approach is used to inform policy and decision-making in health care, especially on how best to allocate limited funds to health interventions and technologies (. . .)"* [29]. It is, undoubtedly, of multidisciplinary nature, involving areas such as clinical medicine, epidemiology, biostatistics, economics, ethics, sociology, systematic reviewing, and meta-analysis [30].   Simply put, HTA brings together scientific evidence and policy-making to assess the overall value of new or existing technology (dependent on its benefits, risks, and costs).

Despite being well developed in the pharmaceutical industry, the HTA of medical devices is still in development and varies, according to the country and the health system implemented. In high-income countries, with developed health systems, HTA processes are used to assess which new or costly technology should be introduced in the health system. However, in low (and middle) income countries, HTA is used to consider which technologies should be included or to perform prioritization when resources are scarce [31]. One important method used in HTA reports and by HTA agencies

is EBM (Evidence-based Medicine) and its pre-existing studies, such as Randomized Clinical Trials (RCTs), comparative and case studies, and expert opinions [32].

Initially, national HTA agencies focused on producing elaborate assessment reports, that could be used by a variety of decision-makers, according to the problem they had in their hands. However, HTA has spread beyond HTA agencies and nowadays hospitals can perform local HTA models, smaller approaches designed to access, for example, the purchase of a new drug for an orphan disease, the introduction of expensive technology in the existing and everyday clinical practice, or a new indication regarding the application of an already existing technology.

Regarding HTA agencies, there is a variety of them, some of which are discriminated bellow [31]:

- The International Network of Agencies for Health Technology Assessment (INAHTA);

- The Health Technology Assessment international society (HTAi);

- The International Information Network on New and Emerging Health Technologies (EuroScan International Network);

- The European network for Health Technology Assessment (EUnetHTA).

HTA can be of various types, ranging from the ones conducted by National HTA agencies (national level HTA) to a regional level, more focused on certain issues, and lastly, local (or hospital level) – the so-called HB-HTA.

Regarding HTA of medical devices, one can observe that, in Europe, there are many heterogeneities, regarding which type of HTA processes are conducted. This variation is not only in the type of HTA but also in the degree of formality associated [33]. Ultimately, this lack of homogeneity is seen as something that needs to be improved: currently, and depending on the country, medical devices are assessed in different ways, in some cases with more and others with less assessment. As a direct consequence, different evaluation pathways and methodologies are adopted and different requirements are demanded, in terms of evidence necessary to gather. This is undoubted, one of the main challenges that the HTA of Medical Devices is facing.

Therefore, and despite the HTA's important role in health technology decision-making, there is still evidence of some fragilities associated. A recent study, published in 2020, concluded that HTA recommendations widely vary throughout different jurisdictions, sometimes for the same health technology (in this specific case, for the same medicine) [34]. Differences in the inclusion and/or evaluation of the evidence or its impact will ultimately affect the final HTA outcomes. This has, of course, direct consequences, since these variations on country jurisdictions may influence the conclusions provided by HTA bodies.

To improve the current decisions associated with health technologies, policymakers have started to gain interest and to adopt HB-HTA. Not only because of the reasons already mentioned but also to use the available resources more efficiently, due to their scarcity in the healthcare system context, to improve both the quality and efficiency of care delivery when limited budgets are used [35].

3.1.1 Hopital-based HTA

HTA can also be conducted in the hospital context, for managerial decisions. This type of HTA is called HB-HTA [33]. HB-HTA started in the mid-'90s, mainly focused in the Northern Europe region, Spain, Italy, Canada, and Australia [48]. According to the literature, the number of hospitals performing HB-HTA is increasing, which calls for the need to improve its quality and efficiency and develop guidelines and tools to support it [35].

HB-HTA consists of performing HTA activities, in this case, tailored for the hospital context regarding multiple types of health technologies. These HB-HTA activities can be "in" and "for" hospitals. While HTA

"for" hospitals is performed by external bodies, such as consultations, temporary contracts, and even freelancer activities and/or projects, HTA "in" hospitals is carried out internal and locally by a team of local hospital professionals, that can be either clinicians or an actual HB-HTA unit [33]. The final products of a HB-HTA consist of HTA assessment reports, technical queries, and quick response services.

HB-HTA can be performed in multiple ways and with different levels of complexity. It can be an actual unit with permanent HTA professionals, working full time or a group of clinicians dedicated to this part-time activity. The team is usually multidisciplinary, comprising clinicians, epidemiologists, public health specialists, and health economists [33].

In Europe, the most common HB-HTA units act internally within hospitals and do not have large dependencies on HTA organizations. Their main focus is to assess medical devices and equipment.

Literature also states that hospitals with an internal HB-HTA typically have better management of the process regarding the adoption of the healthcare technology, and due to that, the assessment period associated is shorter (even though this period is also affected by the health technology in hands).

Due to the increasing interest of HB-HTA, practically worldwide, some groups have been created, namely a HB-HTA Sub-interest Group in 2006 (created within the HTA International Association – HTAi) and more recently, the AdHopHTA, funded by the EU [35].

Needs for adopting HB-HTA

Nowadays, hospitals are facing pressure and challenges regarding being in the front line of innovative technologies. The population is aging and society's expectations are rising regarding the quality and access to healthcare. Simultaneously, innovative technology is constantly being developed, fuelled by new scientific discovery and investigation, allied with overall technological development.

Therefore, it is important to introduce new health technologies, so that life expectancy and quality of life are improved. Nevertheless, one must be aware of the fact that not all health technology provides positive health outcomes and therefore, must be assessed. As already stated, despite being introduced to serve governments, evidence shows that HTA may have a greater impact in the hospital context [33].

Even though HTA reports performed by HTA agencies are usually easily available, hospital managers and clinicians tend to perceive them as slightly loose to their management and clinical practices. The main reasons are the mismatch associated with the topic of focus, content, and timing (some of these assessments are not available at the time the hospital needs them). Besides that, it has also been reported that some clinicians mistrust HTA agencies. Below, some of the reasons to adopt HB-HTA are exhibited [33]:

1. Stable and/or tightening hospital budgets, combined with an increasing influx of new technologies make prioritisation a necessity;

2. HB-HTA provides hospital decision-makers with science-based, multifaceted information and the necessary arguments on which to base the decision and invest in technology;

3. Information from HB-HTA is superior to HTA agencies because:

   - Rapid and timely;
   - Tailored to the hospital settings;
   - Tailored to the requirements of hospital managers.

4. HB-HTA increases the effectiveness of the technologies used in hospitals;

5. HB-HTA improves efficiency in hospital budget management;

6. HB-HTA may lead to improved patient safety.

Steps for adopting a health technology

The process of adopting a certain health technology, within a hospital context consists of a set of steps, explained in the schematic (Figure 3.1) [33].
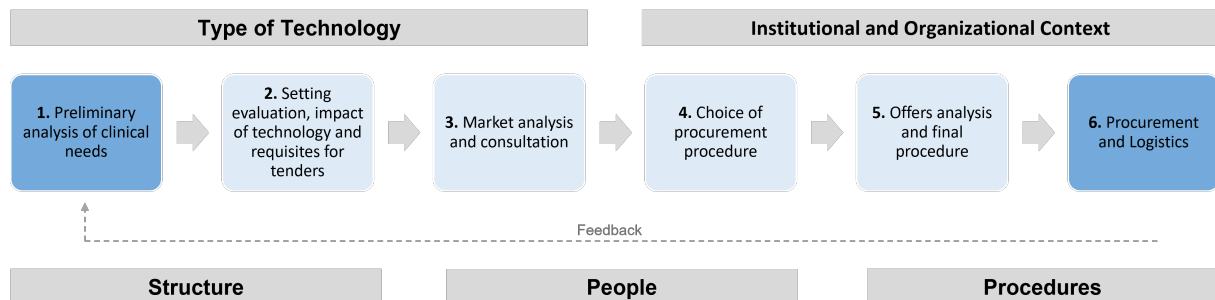
Figure 3.1: Steps for the adoption of a health technology, in hospital context. [adapted from: [33]]

The first step is associated with a preliminary analysis of the clinical needs, whether it is related to the disease burden, available options of medical devices for a certain treatment, number of patients that require specific treatment, among so many others. The second step is associated with setting evaluation, *i.e.*, evaluation of the appropriate level of care in which the technology is intended to be used. The impact of the technology and the definition of requisites are also evaluated. Next, market analysis and consultation are performed, followed by a choice of the procurement procedure (the procedures associated with the purchase of health technology – for example, choice of tenders, negotiations related with the price, approval, and receipt of payment, etc). After that, an analysis of all the offers is performed and a final decision is taken. The final step is related to the procurement and logistics that are associated with the introduction of the specific technology.

It is important to note that these steps can be affected not only by the types and other specific characteristics of the health technology that is being analysed but also by the hospital characteristics (whether organizational or procedural) and the individuals and teams involved in the decision-making process [33].

HB-HTA reports

According to the hospital decision maker's needs, the characteristics of the HB-HTA report will vary, depending also on the health problem that is being evaluated, the evidence quality, and availability of the type of technology and its life cycle.

Based on the literature, there are two types of HB-HTA reports that are the most commonly used in the hospital context: a full HB-HTA report and mini-HTA [33] [36].

Mini-HTA was first developed in Danish hospitals and since then, they are being improved and adapted, worldwide [37]. As one might conclude from its name, it is a short, yet structured assessment regarding both prerequisites and consequences of the use of specific health technology for a predetermined group of patients, hospital level-wise [33]. The typical mini-HTA, usually for the introduction of new technology in hospital and clinical context, is in the form of a questionnaire, with questions concerning both the pros and cons of investing in this new technology, grouped in four main elements: technology, patient, organization, and economy. It is typically retrospective, in the sense that is based on relevant literature review (systematic or not) and opinions provided by experts [33]. Due to its development within a short timeframe, it can be considered as a flexible, adaptable, and dynamic tool, being easily adaptable to the imposed local budgets and policies [38].

A comprehensive (or full) HTA on the other hand, provides an interdisciplinary and systematic

assessment of the health technology (also for a targeted group of patients, in a hospital context). It is a broader assessment, when compared with a mini-HTA since it focuses, not only on the four main aspects of the mini-HTA – technology, patient, organization, and economy – but also on ethical and social aspects included. The data used for this type of assessment is not only primary, (*i.e.*, generated for this purpose), but also secondary, in exhaustive and systematic literature review [33].

Unlike a comprehensive HTA, which usually lasts one year and a half and can last up to three years, a mini-HTA typically lasts months [38]. Therefore, the speed and timing of this assessment are an evident advantage. However, it does not possess the quality and toughness of a comprehensive HTA and, in some more complex cases, a mini-HTA cannot replace an HTA.

Mini-HTA can exist in other versions and can also be combined with other models, such as a combination with MCDA, for instants [36].

Despite this two being the most used, there are other types of HB-HTA, that are considered to be "in-between" mini-HTA and full HB-HTA reports [33]:

- List of technologies for potential disinvestment: report carried out to review hospital drug formulary or hospital medical device list for potential disinvestment;

- Mini-HTA (using clinical trial data or any routinely collected data): prospective report for primary research on clinical efficacy and cost-effectiveness of innovative technologies that have just entered the market (typically medical equipment, medical devices and diagnostic tests);

- Technical input: joint initiative of HTA makers and users (hospital committees) to manage the introduction of a specific health technologies, typically medical devices, used across different clinical department (units);

- Medico-economic analysis: report to assess both the budget and medical impact of a technology (medical devices and drugs);

- Semi-rapid HTA: type of report used to provide evidence-based background information for decision-making. This report is used in hospitals in Finland for all types of health technologies except drugs;

- Rapid systematic review: to provide evidence-based background for hospital decision-making. It includes medical equipment, medical devices, clinical procedures. and drugs, among others;

- Drug assessment: also used to provide evidence-based background for hospital decision-making, but only for the specific case of drugs.

The choice of the report will therefore depend, not only on the type of health technology is being analysed but also on the balance between the need for quality and robustness against the resources and time available [33].

3.1.2 Collaboration between HTA agencies and HB-HTA units

According to the literature, and as already mentioned, the number of HB-HTA initiatives is increasing worldwide. Despite their presence in many countries and regions, HTA agencies do not always provide the best assessments regarding the hospital's needs and requirements. On the other hand, HTA agencies have larger resources and experience to perform high-quality assessments, when compared to local HB-HTA units. That is why the collaboration between these two entities, in a process of joined forces, can be of extreme relevance to hospitals.

Taking European countries as an example, the majority of them only have informal HTA activities between HTA agencies and HB-HTA units, even though formal collaborations are considered necessary and more beneficial [33]. This informal system of collaboration is associated only with ad hoc contacts,

while in the formal system, there are people nominations, followed by a formal process for sharing information, participating, and give feedback. The most common form of collaboration is informal and typically consists of document sharing and training. However, it is also stated that individual expertise and political support are also shared.

As already referred, not all European countries have the same types of collaboration between HTA agencies and HB-HTA units or initiatives. Countries like Norway and Finland have a formal system collaboration between the national and the hospital level. The Basque country (in Spain) and the Italian regions of Lombardy and Lazio also have a formal collaboration, except that, in this case, the collaboration is between regional and hospital level. Cases of informal collaboration also exist, in countries such as Austria, Catalonia (in Spain), Denmark, Switzerland, and Turkey.

However, despite the existence of these collaborations, there are still some barriers that must be faced, the most obvious ones associated with lack of culture and the voluntary nature of using HTA, in the hospital context, and decisions. It has also been reported different expectations concerning timelines and overall methodological quality of HTA reports by the two entities. Another important barrier that was stated is associated with the fact that the overall lack of regulations and other forms of national policies to mandate the use of HTA in a more systematic way, leads hospitals to a rejection path in regards to this collaborations, mainly because of the clinicians' fear of losing their professional autonomy. Adding to this, the high standards imposed for the methodology of national or regional HTA agencies seems also to be a barrier to the involvement of HB-HTA units in these types of collaborations [33]. Nevertheless, some facilitators for these collaborations were also identified, the most important one being the need for a formal and systematised organization for collaboration. However, the need for some informal collaboration (in the form of informal contacts between individuals, for example) is never ruled out. [33].

3.1.3   Gap between clinical research and clinical practice

Throughout the years, it has become increasingly evident that clinical practice cannot keep up with the pace at which current clinical evidence is being generated [39].

This fact can be justified by an overall lack of appropriate systems, frames, and/or strategies that would easily bridge the gap between scientist/researcher and practitioner, often controlled by policymakers [40]. What is stated is that there is little contact between these two distinct worlds [32]. The flagrant differences between these cultures compromise the bridging of these two worlds and therefore increases the gap between them. Despite this fact, research waste is also reported, whether because the studies were not relevant neither to clinicians nor patients, the studies were improperly conducted (with biases associated), or because there was not full access to the studies. Some even report the low quality of much of the research published in the literature [32]. All these factors limit the passage of clinical evidence to practice (Figure 3.2).

One possible way to diffuse the information to the clinical practice is by issuing Clinical Guidelines. Its main goal is to improve the overall quality of the healthcare practice and to minimize possible variations in healthcare practice that could increase patient outcomes [32]. Besides that, it is also necessary to break barriers between scientist and clinician to promote dialogue and promote training in evidence-based practice among clinicians, while developing methods to easily translate quality research into clinical practice and clinical insight into the empirical study [41].

The existence of a gap also exists in the relationship between scientists and policymakers. Whether it is because there is a lack of time, support, or incentives for scientists or because it is difficult to solve, in a fast way, the demand for specific information. The vast majority of the scientific studies do not address policy-related issues and other forms of knowledge required for the policy process [40]. Besides that, it is also stated that policymakers also lack the skills to understand the available scientific evidence [40].

As one can conclude, it is a challenge to transfer evidence from medical research and introduce it to clinical practice. This gap is prominent, and according to the literature, it has become larger. Not only that but the medical possibilities and opportunities also grow at a more rapid pace than the available health care budgets, which leads to the necessity of adopting not only effective but also cost-effective technologies, with the purpose of not overwhelming health care systems and also to discard the ones that are not effective [32].

Questions relevant to clinicians and patients?	Appropriate design and methods?	Accessible full publication?	Unbiased and usable report?
•Low priority questions addressed •Important outcomes not assessed •Clinicians and patients not involved in setting research agendas	•Over 50% of studies designed without reference to systematic reviews of existing evidence •Over 50% of studies fail to take adequate steps to reduce biases (*e.g.* unconcealed treatment allocation)	•Over 50% of studies never published in full •Biased under-reporting of studies with disappointing results	•Over 30% of trial interventions not sufficiently described •Over 50% of planned study outcomes not reported •Most new research not interpreted in the context of systematic assessment of other relevant evidence

Research waste

Figure 3.2: Stages of waste in the production and reporting of research evidence relevant to clinicians and patients. [adapted from: [32]]

3.1.4 HTA of biomarkers

There are multiple ways to evaluate new and emerging biomarkers. Since biomarkers are medical devices they can be evaluated through a HTA process.

The type of HTA process used, however, depends on which phase of the discovery and development pipeline the biomarker is located. For example, in an early evaluation, in the developmental phase of a new biomarker test, the focus relies on which biomarkers have sufficient potential, *i.e.*, which ones are effective, and on how cost-effective they are. Thus, economic evaluations of the biomarkers should be conducted. However, in a later stage of the evaluation, the focus shifts and relies on the biomarker's risk prediction and economic evaluations in a clinical setting, *i.e.*, already being used in the hospital context. Ahead, a focus on both effectiveness and economic evaluations is presented.

Effectiveness evaluations

To evaluate the effectiveness of a biomarker, the standard, classical way is to assess both its sensitivity and specificity. Sensitivity measures the ability that the biomarker presents to correctly identify a positive case, while specificity refers to the biomarker's ability to correctly identify a negative case [42]. Together they form the base for several statistical methods for characterizing the degree of improved predictive power.

One of the most common examples includes calculating the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve of a new, emerging biomarker and compare it to already qualified ones, to assess their clinical added value. In fact, if the difference between the AUC of a new biomarker and the AUC of a qualified one is positive, the emerging biomarker has clinical added value; if the difference is null, then there is no difference between the AUCs of both biomarkers, meaning that the new biomarker does not have clinical added value; lastly, if the difference is negative,

the new biomarker presents less value than the current qualified ones [43] [44].

The ROC curve shows the trade-off between the sensitivity and 1-specificity, given by the True Positive Rate (TPR) in the y-axis and the False Positive Rate (FPR) in the x-axis, respectively. These rates are given by the following equations:

$$TPR = \frac{False\ Positives}{False\ Positives + True\ Negatives} \tag{3.1}$$

$$FPR = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3.2}$$

Thus, the AUC measures the two-dimensional area under the ROC curve. Ideally, a perfect biomarker would behave as a perfect classifier, *i.e.*, with sensitivity and specificity of 100%, meaning that it would result in no false negatives nor false positives (meaning its predictions were always correct), and therefore with an AUC of 1. Biomarkers with an AUC equal to or below 0.5 behave as a random classifier and are considered irrelevant, with no predictive value. Consequently, the higher the biomarker's AUC is (between 0.5 and 1), the better the biomarker's performance [43] [44].

Other examples consist of the Net Reclassification Improvement (NRI) [44], and Decision Curve Analysis (DCA) over a plausible range of cut-off for predictions [42].

Economic evaluations

Despite the evaluation of the biomarker's clinical utility, *i.e.*, if it is useful in patient care and clinical decisions, it is also relevant to determine its economic benefit. Evidence suggests that studies to assess the economic benefit of biomarkers for laboratory tests are not as common as other studies for more expensive treatments or interventions, due to a smaller number of randomized controlled trials for biomarkers and also the small healthcare budget fraction that is used for this type of technology [45].

Some of the economic evaluations that can be performed are [45]:

- **Cost-minimization**: to determine, in a group of tests that will produce the same outcome, which is the least expensive;

- **Cost-effectiveness analysis**: to assess, most efficient way to use a fixed amount of resources to obtain the largest effect (*e.g.*, using the life year natural unit);

- **Cost-benefit analysis**: to compare the benefits and the costs of the test. However, it is sometimes difficult to assign a monetary value to the benefit, which requires equating a monetary value to a year of life;

- **Cost-utility analysis**: to estimate the ratio between the cost of the test and the benefit it produces in the number of years gained. These types of tests are the most frequently used.

It is important to note that in the case of emerging biomarkers there is still little evidence available, which makes it very challenging to conduct these types of studies. Nevertheless, it is extremely important to conduct these types of studies to determine if a biomarker will make economic sense, both from the payer's and the society's perspective [45].

3.1.5   MCDA in HTA

Besides the more traditional effectiveness and economic evaluations, there are new assessment approaches which use different evaluation dimensions. As already mentioned, there is a high number of biomarkers that enter the biomarker discovery and development pipeline, which makes it difficult to evaluate them with the available monetary and time resources. Thus, and in order to select the most promising ones to be used in clinical practise, the use of a tool such as MCDA is of extreme relevance.

In fact, MCDA has been considered extremely useful in HTA, due to its alignment with value-based healthcare, its transparency and ability to account for the different stakeholders values and preferences, its contribution to aid decision-makers to understand the technology value, among many others [46].

There are several examples regarding the use of MCDA in healthcare, including the assessment of health technologies regarding its benefits, risks and costs. Typically, decision-makers, in cases of multidimensional healthcare problems, are faced with very complex problems and so, in order to simplify them, they tend to opt for more intuitive or heuristic approaches. As a direct consequence, important information can be neglected: usually, evidence from economic evaluations tends to be the one that is more used, with social values and context being under-used, little analysed and sometimes even excluded from the decision-making process. This can ultimately lead to decision-makers not being adequately equipped to make rational and well-informed decisions and diminished credibility related with the decision outcomes [47]. That is why the use of specific frameworks is so important, since it brings a more detailed, structured and transparent way of assessing and choosing healthcare technologies, leading to better and trustworthy decisions and solutions.

One example is EVIDEM, a framework developed for decision-making, associated with components of MCDA [48]. Simply put, the EVIDEM framework is based on three distinct principles: (i) support the deliberative process by deciphering and quantifying scientific judgement, when possible and value judgement; (ii) improve access to relevant evidence, using a collaborative structure; (iii) enhance communication of decisions using transparent tools [48]. Thus, this framework provides a comprehensive and transparent structure grounded in global standards and local needs. It can be used on a variety of healthcare decisions, namely to generate data on quality of evidence, adapt framework to existing decision-making processes, develop methodology to generate data tailored to critical data needs or define explicit needs of decision-makers, among so many others [48].

Another example is the Advance Value Framework, a methodology also based on MCDA and adapted for the needs of HTA [49]. It follows a process using a top-down "value-focused thinking" approach, that involves literature reviews and expert consultations [47]. This framework uses model building and selection of criteria to develop a generic value-based model, in the form of a value tree, that can be used in multiple types of healthcare technologies, not only medicines/drugs and medical devices, but for other types of healthcare interventions as well. By combining the MCDA modelling techniques with the construction of value preferences with the value tree, this Advance Value Framework allows to measure the value of healthcare technologies, enabling the assessment of new technologies, in a transparent and very structured way [47].

Lastly, another example of a MCDA method is MACBETH [50] which was used for the development of this thesis. In the following chapter (Chapter 4) a detailed explanation of this MCDA method will be conducted as well as a methodology using MACBETH is proposed and described.

3.2 Breast Cancer

Breast cancer consists of cancer that is formed in the breast tissues, typically in the ducts (*i.e.*, tubes that transport milk to the nipple) or in the lobules (*i.e.*, glands that produce the milk) [51]. It is considered the most common carcinoma in women.

Breast cancer can be categorized as non-invasive or invasive. The non-invasive breast cancer also called *in situ*, corresponds to a pre-malignant lesion. Invasive breast cancer corresponds to the cases in which cancer has already spread outside the place in which it initially developed: the ducts (invasive ductal breast cancer) or lobules (invasive lobular breast cancer) [51].

Despite being referred to as a single disease, breast cancer is very heterogeneous, with a large spectrum of different diseases. Different breast cancer subtypes have varied morphological

characteristics and distinct clinical outcomes.

In the following sections, a review of the current evidence regarding breast cancer will be presented.

3.2.1 Epidemiology

Demographics, incidence and mortality

Breast cancer is considered the most frequent malignancy in women worldwide, with a lifetime risk of 1 in 10, extremely high when compared with 1 in 833, in the case of men, which can be considered rare. In 2018 only, around 2.1 million women were newly diagnosed [52]. Despite the incidence steady rise throughout the years, due to overall population growth and aging of the population, the mortality remains relatively stable with over 70% of all breast cancer cases being treated, translating the technological improvements for early diagnosis, treatment options, and prevention screenings [53].

Breast cancer incidence also differs across regions and countries [52]. This is a direct reflection of not only risk factors but also the availability and utility of screening mechanisms and programs. Generally speaking, in developed countries there are higher numbers of mammography machines that are used broadly than in developing countries, increasing inevitably the incidence of breast cancer cases in the first region. Nonetheless, developing countries also have high numbers of breast cancer cases. However, they are diagnosed at a later stage of the process, when the disease is already advanced, leading to worse health outcomes and increased mortality rates. That is why developing countries experience low breast cancer incidence, but higher mortality associated, due to late diagnostic and limited access to proper treatment, if any [52]. Additionally, the biology of breast cancer differs from ethnicity, also leading to distinct mortality rates [52].

Genetic predisposition

Regarding the genetic predisposition of the disease, it is considered that approximately 10% of all breast cancers are genetically caused and inherited (and associated with family history), despite not being homogeneously spread across races and different ethnicities (as already mentioned above) [52].

A higher risk of breast cancer (approximately 20 to 25% of breast cancer risk) has been associated with mutations in two distinct high-penetrance tumour suppressor genes – BRCA1 and BRCA2 - despite other genes may also be associated (corresponding to approximately 2% of breast cancer causes, such as STK11, ATM, PALB2, PTEN, CHECK2, P53, and CDH1 genes) [52] [54].

Risk factors

As already mentioned, the epidemiology pattern of breast cancer differs across countries (due to distinct breast cancer management), ethnicity, and genetic predisposition factors. However, there are other risk factors also associated with the development of breast cancer, such as environmental ones or others associated with lifestyle. Nonetheless, it is important to note that, having some risk factors increases the risk of cancer development, but it does not necessarily imply the development of breast cancer [51]. Some of the most important risk factors are associated with female sex, increasing age, and having fewer children [51]. Other risk factors include genetic predisposition (with a family history, or mutations in certain genes), exposure to oestrogens (*e.g.* hormonal contraception) [52]), exposure to ionising radiation, or a history of atypical hyperplasia (abnormal and non-carcinogenic cell growth in the breast tissue within the ducts and/or lobules of the breast) [51].

It is also estimated that approximately 20% of all breast cancers worldwide are associated with modifiable risk factors, such as obesity, physical inactivity, and alcohol use. These risk factors are easier to prevent than some of the others mentioned, simply by adopting a healthier lifestyle.

3.2.2 Established classifications of breast cancer

To have optimal patient management, it is necessary to classify the type of breast cancer, not only for a precise prognostic but also to plan for effective therapy. Bellow, breast cancer classifications are presented and discussed, based on the Molecular subtypes, TNM staging system, and Tumour grade.

Molecular subtypes of breast cancer

Based on the surrogate intrinsic classification, associated with histology and immunohistochemistry expression of four biomarkers (Estrogen Receptor (ER), Progesterone Receptor (PR), HER2, and Ki-67), breast cancers can be divided into five different subtypes, presented in Table 3.1.

Table 3.1: Breast cancer subtypes. [based on: [52], [55]]

Subtype	Frequency	Features	Prognosis	Recommended therapy
Luminal A	- 30-40% of all invasive BC	ER+ PR+ ($>$20%) HER2- Ki-67 low ($<$14%)	- Generally good; frequent relapse after a long disease-free interval - Favourable prognostic (G1 or G2)	Endocrine therapy alone in the majority of the cases
Luminal B-like (HER2-)	- 20-30% of all invasive BC	ER+ PR +/- HER2- Ki-67 high ($\geq$14%)	- Lower than Luminal A tumours but better than HER2+ tumours; PR- cases are more aggressive - Unfavourable prognostic (G2 or G3)	Chemotherapy followed by endocrine therapy for the majority of the cases
Luminal B-like (HER2+)	- 15% of all invasive BC; - Around half of all HER2+ invasive BC	ER+ PR +/- HER2+ Ki-67 high ($\geq$14%)	- Lower than other luminal tumours; PR-cases are more aggressive - Unfavourable prognostic (G2 or G3)	Chemotherapy and anti-HER2 therapy followed by endocrine therapy for all patients
HER2-enriched	- 15% of all invasive BC; - Around half of all HER2+ invasive BC	ER- PR- HER2+ Ki-67 high ($\geq$14%)	- Aggressive clinical behaviour, but good response to target therapy against HER2; - Unfavourable prognostic (usually G3)	Chemotherapy and anti-HER2 therapy
Triple Negative	- 15% of all invasive BC	ER- PR- HER2- Ki-67 high ($\geq$14%)	- Variable, but generally aggressive; - Unfavourable prognostic (usually G3)	Chemotherapy

TNM staging system

The most commonly used system is the TNM staging system, published by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) [52]. This classification uses a combination of three variables: tumour size at the primary site (T), lymph node involvement (whether cancer has spread to lymph nodes) (N), and spread to distant metastatic sites (M) [51] [52].

T, N, and M are combined to create different breast cancer stages ranging from stage 0 to IV, the last one being the most advanced case [56]. Stage 0 is considered non-invasive since it is only found inside the ducts or the lobules and it has not spread to the surrounding breast tissue nor the lymph nodes (N0) nor other distant sites (M0). Stages I, II, and III, on the other hand, correspond to invasive breast cancers, where the tumour has grown outside the ducts, lobules, and breast skin, and it can be in the axillary lymph nodes. Lastly, on stage IV, the cancer is metastatic, meaning that it has already spread to distant sites and can also be found in axillary lymph nodes [57].

Table 3.2: Breast cancer TNM staging system. [adapted from: [51]]

Stage 0. Non-invasive tumour confined to the breast (T1N0M0)	
Stage I. Tumour is small and confined to breast tissue or with evidence of cancer in lymph nodes close to the breast	
IA	• The tumour is no bigger than 20mm in diameter and is confined to the breast (T1N0M0)
IB	• There is no evidence of a primary tumour (T0) or the tumour is no bigger than 20mm in diameter (T1), but micrometastases (no bigger than 2mm) are present in the axillary lymph node(s); lymph nodes are movable (N1mi); no distant metastases are present (M0)
Stage II. Tumour is in the breast or in the nearby lymph nodes, or both	
IIA	• There is no evidence of a primary tumour (T0) or the tumour is no bigger than 20mm in diameter (T1); metastases are present in the lymph nodes(s) and lymph nodes are movable (N1); no distant metastases are present (M0); • The tumour is larger than 20mm but no bigger than 50mm in diameter (T2) and is confined to the breast (N0); no distant metastases are present (M0)
IIB	• The tumour is larger than 20mm but no bigger than 50mm in diameter (T2); metastases are present in the axillary lymph node(s) and lymph nodes are movable (N1); no distant metastases are present (M0); • The tumour is bigger than 50mm in diameter (T3) and is confined to the breast (N0); no distant metastases are present (M0)
Stage III. Tumour has spread from the breast to lymph nodes close to the breast, to the skin of the breast or to the chest wall	
IIIA	• There is no evidence of primary tumour (T0), the tumour is no bigger than 20mm in diameter (T1), the tumour is larger than 20mm but no bigger than 50mm in diameter (T2), the tumour is larger than 50mm in diameter (T3); metastases are present in the axillary lymph node(s) and lymph nodes are fixed or matted (N2); no distant metastases are present (M0) • The tumour is larger than 50mm in diameter (T3); metastases are present in the axillary lymph node(s) and lymph nodes are movable (N1); no distant metastases are present (M0)
IIIB	• The tumour (of any size) has extended to the chest wall and/or skin (T4); lumph nodes are not involved (N0) or metastases are present in the axillary lymph node(s) and lymph nodes are movable (N1) orlymph nodes are fixed or matted (N2); no distant metastasis are present (M0)
IIIC	• The tumour of any stage (any T); metastases are present in the axillary lymph nodes(s), in supravicular lymph node(s) or internal mammary lymph nodes(s) (N2 or N3); no distant metastases are present (M0)
Stage IV. The tumour has spread to other areas of the body (any T, any N, M1)	

Tumour grade

The assessment of the histological grade is made based on the Elston- and Ellis-modified Scarff-Bloom-Richardson System. This system is based on three distinct tumour features, and each one of the three features is scored, from 1 to 3 [58]:

- Proportion of cancer cells that are in tubule formation;

- Variation of nuclear size and shape between the cells;

- Number of mitoses, *i.e.*, cell divisions.

A final grade, ranging from 3 to 9, is attributed by adding the individual scores of each feature [52]. The total score is categorized into three grades: G1, G2, or G3. G1 corresponds to a low grade and well-differentiated tumour (score 3-5); G2 represents a moderately differentiated tumour (score 6-7) and finally, G3 indicates a high grade and poorly differentiated tumour (score 8-9). Thus, generally speaking, G1 is considered to be the one associated with the best prognostic, while G3 with the worst one [56] [58]. This classification is an important prognostic factor, especially in guiding therapy in early breast cancer cases.

3.2.3 Biomarkers in breast cancer

Biomarkers for breast cancer play an important role, managing breast cancer patients and are of extreme relevance in the context of therapeutic decision-making [52]. According to the type of breast cancer tumour, determined by the biomarkers present, the therapy can be personalized and adapted. Algorithms have already been implemented, to improve the therapy given to the patients, that can have distinct subtypes of this disease. Different subtypes of breast cancer are originated from the different expressions of biomarkers. Therefore, the treatment is personalized, according to the subtype and current stage of the breast cancer present.

There are multiple biomarkers available and already validated to be used in breast cancer therapy. For all patients with diagnosed invasive breast cancer (for primary, recurrent, and metastatic breast cancer [59]), it is mandatory to determine ER, PR, and HER2 biomarker status, in agreement with the already published guidelines.

Ki-67 is another biomarker that is widely used to determine proliferation and to predict chemosensitivity. It is a protein found in cells when they are dividing, but not when they are at rest. Thus, if there is a high level of Ki-67 in cells, this indicates that those cells are growing very rapidly, translating the existence of a tumour [51]. However, it is only relevant for ER-positive and HER2-negative breast cancers [52]. Thus, there is not a consensus regarding this biomarker and the Ki-67 determination is neither standardized nor generally recommended. Nevertheless, by some authors, this biomarker is considered to classify distinct breast cancer subtypes. By doing so, biomarkers can indicate the prognosis and can help doctors to determine which treatments should be considered for each breast cancer subtype [51].

As already referred, the status of ER, PR, and HER2 in invasive breast cancer patients is mandatory, and these biomarkers have been recognized by international guidelines [52]. Apart from ER, PR, and HER2, recommendations for the clinical use of other breast cancer biomarkers vary according to the recommendations of distinct expert panels (ASCO, NCCN, ESMO, St Gallen Consensus Panel). Nevertheless, the following biomarkers presented in Table 3.3 have all been validated for therapy decision-making, with strong levels of evidence associated with its clinical use [52].

Table 3.3: Biomarkers validated for therapy decision-making [adapted from: [52]]. (CEP17- chromosome enumeration probe 17, IHC- immunohistochemistry, ISH- *in situ* hybridization, LOE- level of evidence, NGS- next-generation sequencing; PARP- poly(ADP-ribose) polymerase, RT-PCR- PCR with reverse transcription).

Biomarker	Method and threshold	Use	LOE
ER	IHC: positive if ≥1%	● Essential for the characterization of the IHC luminal group ● Poor prognostic marker if negative ● Predictive marker for endocrine treatment ● Mandatory for endocrine treatment prescription	I
PR	IHC: positive if ≥1%	● If negative, tumour classified as IHC luminal B ● Strong poor prognostic marker if negative ● Predictive marker for endocrine treatment	I
HER2+	● IHC: positive if >10% complete membrane staining ● Single-prone ISH; positive if HER2 ≥6 copies ● Dual-probe ISH; positive if HER2 and CEP17 ≥2 and HER2 ≥4 copies, or HER2 and CEP17 <2 and HER ≥6 copies	● Essential to characterize HER2-enriched (ER-) disease and luminal B, HER2+ ● Prognostic marker ● Predictive marker for anti-HER2 treatment ● Mandatory for anti-HER2 therapy	I (IHC and ISH)
Ki-67	IHC: no final consensus on cut-off value but values <10% are considered low and >30% are considered high	Absence of international consensus for scoring and threshold	I
		Prognostic value in ER+, HER2- tumours (primary tumours and post-neoadjuvant tumour residues)	I
		Absence of prognostic value in HER2+ disease or TNBC	I
		Predictive of response to neoadjuvant chemotherapy	Expert opinion
		If elevated, chemotherapy is often prescribed in ER+, HER2- tumours	Expert opinion
		Part of the IHC definition of luminal tumours whereby when Ki67 is low, luminal A tumour likely and when Ki67 high, luminal B tumour likely	Expert opinion
Intrinsic subtypes	Gene expression profile, N-counter technology	Prognostic	II and III
		Predictive; different responses to neoadjuvant chemotherapy and anti-HER2 therapy according to subtype	I
1st Generation signatures (MammaPrint and Dx)	Gene expression profile, RT-PCR	● Prognostic for ER+, HER2-tumours (with 0-3 involved lymph nodes) ● Chemotherapy is indicated if high risk or high score	Ia
2nd Generation signatures (Prosigna and Endopredict)	N-counter technology, RT-PCR	● Prognostic for ER+, HER2-tumours (with 0-3 involved lymph nodes), include T size and N status in their final score ● Chemotherapy is indicated if high risk or high score	Ib
PIK3CA mutations	Mutations detected by PCR or NGS in exons 9 and 20 from cancer biopsy specimen or liquid biopsies	Predictive marker for specific PIK3CA inhibitors (such as alpelisib) in luminal A and luminal B metastatic breast cancer	Ia
Germline BRCA mutation	NGS on blood lymphocytes or on tumour cells	● Predictive marker for PARP inhibitors in metastatic breast cancer (evidence-based for HER2- disease) ● Germline mutations imply family counselling ● Predictive impact of somatic mutations is under evaluation	Ia
PD-L1	IHC; positive if expression in immune cells ≥1% in tumour specimens	Predictive for immunotherapy with atezolizumab combined with nab-paclitaxel in TNBC	Ia

26

3.2.4 Treatment

Currently, the treatment options for breast cancer consist of surgery, radiotherapy, chemotherapy, systemic therapy, and other more non-conventional treatments [51]. These multiple treatment options for breast cancer will ultimately depend on the size, location, the number of tumours, and the cancer pathology associated with the tumour (*i.e.*, its subtype, grade, and the presence of biomarkers). Other factors, such as age and overall health of the patient should also be considered when choosing the best treatment method.

To aid the medical staff involved in the decision for the patient's treatment option (that involved a multidisciplinary team, comprising doctors from different specialties, nurses, and other health professionals), algorithms have been developed to facilitate the decision process for the therapeutic procedure, both for early and late-stage breast cancers (Figure 3.3a and Figure 3.3b, respectively) [51] [52].

(a) Early invasive breast cancer.   (b) Advanced breast cancer.
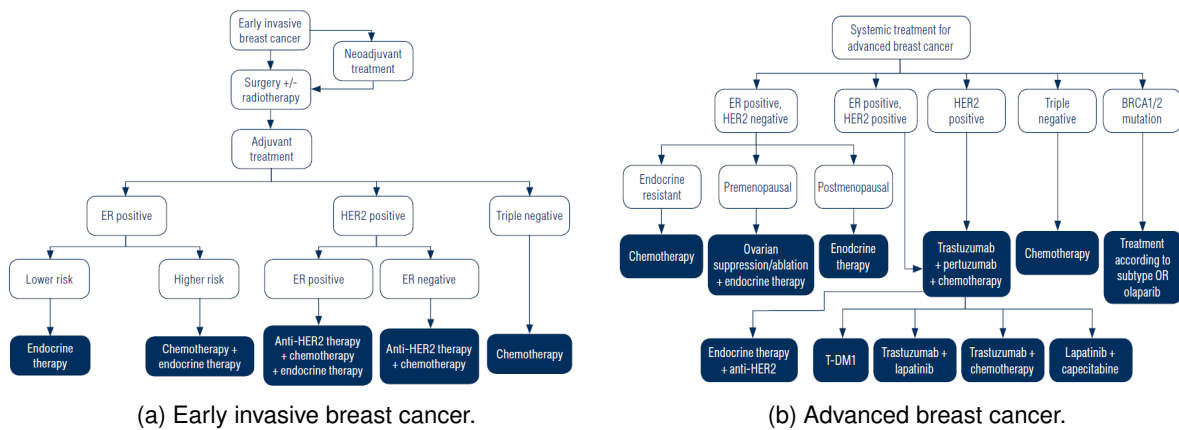
Figure 3.3: Current algorithms used for treatment approached of early invasive breast cancer (a) and advanced breast cancer (b), according to its subtype. [adapted from: [51]]

3.2.5 The case of HER2+ breast cancer

HER2 and breast cancer

HER2 is a transmembrane glycoprotein, coded by the ERBB2 gene (commonly known as HER2/neu) [60]. It is part of the Human Epidermal Growth Factor (HER) family receptors that comprises other 3 members: HER1 (or EGFR), HER3, and HER4. These receptors are associated with pathogenesis in a variety of cancers, besides breast cancer [60]. Additionally, they are also responsible for the regulation of cell functions, namely growth, survival, proliferation, and differentiation.

The amplification or overexpression of HER2 occurs in approximately 15-30% of all breast cancers (around 1 in 5 women diagnosed with breast cancer have HER2+ breast cancer), also being expressed in other cancer types, namely gastric, esophageal, ovarian, endometrial, lung and bladder carcinomas [60]. Its presence is associated with facilitating excessive and/or uncontrolled cell growth as well as tumorigenesis.

HER2+ cancer cells possess around 2 million HER2 proteins on their surface, which is around 100 times more than a normal cell (usually around 20.000 units), leading to tumour cells growing and dividing more rapidly [61]. Besides this fact, dimerization (*i.e.*, the pairing of HER proteins) is a vital process in

signalling pathways that ultimately lead to cancer cell growth and encourage the cell to multiply. To do so, HER2 proteins can pair with other HER2 proteins or with other members of the HER family receptors.

As a result, HER2+ breast cancer tends to be more aggressive than the HER2- subtypes (especially Luminal A-like and Luminal B-like (HER2-) subtypes), correlating with poorer prognosis and unfavourable tumour characteristics, such as tumour size, high nuclear grade, and high proliferation index [61].

HER2 is considered an important and reliable biomarker for the diagnostic, prognostic, and prediction of drug/therapy response in HER2+ breast cancer. In fact, it is currently the only biomarker approved to guide HER2-targeted therapy [62].

Testing and targeting for HER2+ breast cancer

To test for HER2 positivity in breast cancer tumours, it is necessary to test the cancer biopsies, removed from the patients. This is performed typically in the Clinical Pathology Service laboratory (in case it is done in a hospital). There, the biopsy samples are treated, stained, and analysed under the microscope.

Currently, two methods to test for HER2 positivity in breast cancers are recommended by ASCO and CAP organizations, namely [63]:

- **Immunohistochemistry (IHC)**: used to measure the presence of HER2 receptors. The result is given from a score ranging from 0 to 3+. If the sample has a score of 0 or 1+, it is considered negative; if 3+, it is positive; if the score is 2+, the result is considered to be borderline and further tests must be performed [63];

- *In situ* **hybridization (ISH)**: used to measure the HER2/CEP17 ratio (*i.e.*, the ratio between the number of copies of the HER2 gene and the number of copies of chromosome 17 (where the HER2 gene is located). A ratio is performed to prevent cases of pseudo-amplification [63]. This can be performed using a:

  - Single-signal (HER2 gene) assay (single-probe ISH);
  - Dual-signal (HER2 gene) assay (dual-probe ISH).

The full optimal algorithm for HER2 testing by the ASCO/CAP recommendations can be consulted in Appendix A. Ahead, a focus on the algorithms for HER2+ breast cancer is represented (Figure 3.4).
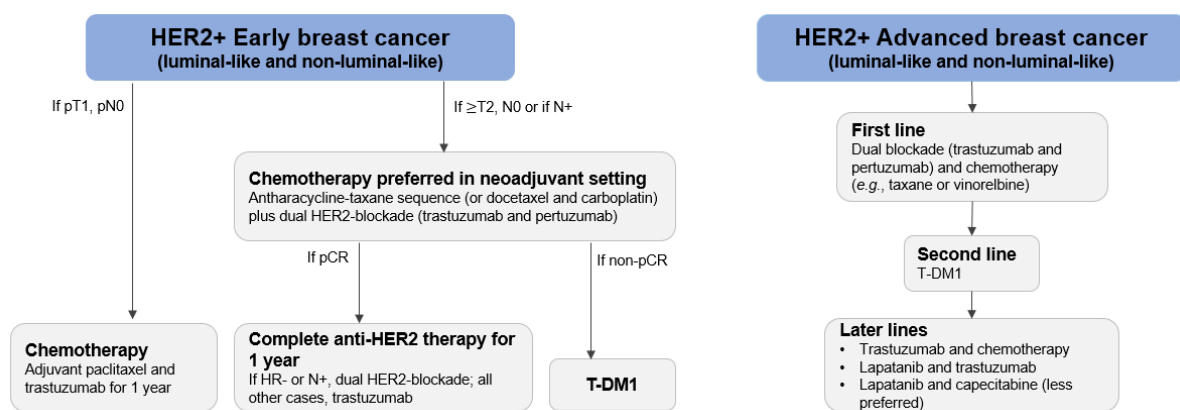
Figure 3.4: Algorithms for HER2+ early breast cancer and HER2+ advanced breast cancer management (both luminal and non-luminal-like). [adapted from: [52]].

Resistance to treatment in HER2+ breast cancer

As already referred, HER2+ breast cancer consists of a very aggressive subtype of breast cancer, both clinically and biologically. Due to new and improved anti-HER2 targeted agents, the patients' prognostic

has increased considerably [64]. Nevertheless, and despite the improved prognostic of HER2+ breast cancer patients, with the use of targeted therapies for the HER2 receptor, more than 60% of patients experiment progression during the first year *i.e.*, resistance to treatment [65]. Many causes of this intrinsic resistance are reported in the literature, namely genetic modifications of HER2 receptor; alterations in the intracellular signalling pathways; defects in cell arrest and apoptosis; defects in the antibody-dependent cellular cytotoxicity (ADCC), among others [65].

To date, in clinical practice, the use of anti-HER2 agents for therapy is based on the HER2 status and Hormone Receptor (HR). [52] [64]. However, evidence from clinical trials also suggests that HER2+ breast cancer is clinically heterogeneous (not only HR- (HER2-enriched) or HR+ (Luminal B-like (HER2+)), since some cases are resistant to treatment, while others are not. Therefore, the current challenge has been to identify the portion of patients that can be treated only with anti-HER2 therapies, without the toxicity of chemotherapy [64]. Thus, despite all the research and the identification of several biomarkers, no treatment for HER2+ disease is indicated based on specific biomarkers. Adding to this fact, it is considered that the HER2 status remains insufficient to clarify the heterogeneous therapeutic outcomes [64] [66].

Consequently, although HER2+ breast cancer is currently dichotomized in HR- (HER2-enriched) and HR+ (Luminal B-like (HER2+)), evidence suggests that many other sources of biological heterogeneity are present [64].

According to the literature, some sources of heterogeneity in HER2+ breast cancer are [64]:

- **Genetic expression**, that includes, for example, the levels of HER2/neu mRNA;

- **DNA mutations**, that includes mutations in specific genes such as PIK3CA or HER2/neu;

- **Immune microenvironment**, *i.e.*, the cells, molecules, and blood vessels that surround and feed the tumour cells. TILs, the Programmed death-ligand 1 (PD-L1) molecules, and Fragment crystallizable-gamma receptors (FcγRs) (expressed in a variety of immune cells) are examples considered to be a source of biological heterogeneity.

Understanding these sources of heterogeneity can eventually lead to new biomarkers for HER2+ breast cancer, expanding the current treatments and, gradually optimize the situation of HER2+ breast cancer patients regarding prognostic and treatment [64]. With a larger number of biomarkers for this breast cancer subtype, it might be possible to decrease the current percentage of patients that are resistant to the therapy and to increase the patient's life expectancy, with more personalized therapies, adjusted to their conditions.

Having this in mind, a Socio-technical approach based on MACBETH will be presented in the following chapter: Chapter 4, where the multicriteria decision analysis model developed to evaluate emerging biomarkers for HER2+ breast cancer subtype will be presented.

Chapter 4

Methodology

In this chapter, the proposed methodology will be explained. First, a summary of the socio-technical approach that was applied in this thesis is presented, including a theoretical exploration of the MACBETH method, steps and concepts associated. Then, the specific steps required to perform the evaluation of biomarkers for HER2+ breast cancer are presented.

4.1   Overview of the proposed socio-technical approach

MCDA consists of a general framework that supports complex decision-making situations, with a large variety of objectives (that are typically conflicting), valued differently by the stakeholders' groups and/or decision-makers, that typically present different opinions. In MCDA it is possible to openly address the distinct problems and opinions that may arise, allowing for a more transparent, reliable and straightforward process. [67]. It can group several concerns in a single model, in a comprehensive and flexible way, which in its construction includes different judgements. This type of model allows the simplification of a complex problem in several problems with smaller dimensions, which are analysed independently and then integrated into a global analysis.

The proposed methodology, based on MCDA, follows a socio-technical approach, in which the design of assessment techniques is performed in combination with social processes. Thus, this process involves, not only, a more technical component, associated with the use of MACBETH to model the selection of emerging biomarkers, but also a social component, to obtain the experts' opinions that will be used for the modelling. This type of approach is very advantageous since it "*improves communication (...), develops shared understanding (...) and generates a sense of common purpose*"; besides "*the real payoff is in smarter decisions that increase the value created by the available resource*" [68].

It is expected that the constructed socio-technical approach becomes an important contribute in the field of HTA, especially for biomarkers, providing guidance to evaluate and determine which are the most promising emerging biomarkers that can be used in clinical practise. Besides this, the model can also be reused, since it can be adapted to other types of medical devices. The different stages of the socio-technical approach used in this thesis will be described in further detail throughout this chapter.

The socio-technical approach (Figure 4.1) used can be divided into four distinct stages: evidence synthesis, model structuring, model construction, and model testing and validation. Each one of these phases can also be divided into its technical and social components.

The technical component includes the definition of the problem, selection of the biomarkers, and the use of MACBETH. The MACBETH approach was chosen for this thesis since it uses qualitative judgements of difference in attractiveness not only to generate the value scores for the options but also

to weight the criteria [50]. By doing so, this tool becomes very intuitive for the decision-maker when compared with other approaches and models that can, oftentimes, be more confusing. The MACBETH approach is used in multiple contexts, not only in the field of health and health technology such as this thesis, but also in other sectors namely energy, military and industry sectors, among others [69]. In the following subsection, a more in-depth explanation of this approach will be presented, since it is important to clarify it before the detailed analysis of the methodology steps.

On the other hand, the social component uses input provided by the participants (in this case, the healthcare professionals from HESE) to build the multicriteria model. In this way, the social component acts as a supporter of the technical component, providing in the specific case of this thesis, individual interviews with experts and decision-makers, a Workshop, and a Decision Conference.

Using this methodology, it is expected to construct a decision support model to assist in the evaluation of biomarkers for the specific case of HER2+ breast cancer, that can later be used for *in vitro* tests. Due to its universality in the world of biomarkers, any biomarker can be assessed using this tool. This can be extremely advantageous since it will be possible to use this multicriteria evaluation model for any new emerging biomarkers that might appear in the literature with the potential to be used in clinical practice.

The first phase in this methodology was to synthesise the available evidence regarding emerging biomarkers for HER2+ breast cancer and to select the most promising ones, based on the information available in the literature. To do so, a rapid review on this area was performed (using academic research databases and other search engines), to select a list of emerging biomarkers for HER2+ breast cancer. Parallel to this step, the results of a Web-Delphi process from the MEDI-VALUE project, prior to the development of this thesis, were used, which served as a starting point for the model.

Next, the evaluation model was structured, *i.e.*, the decision problem was defined, which allowed the selection of the biomarkers to be evaluated. This was possible due to the input provided by experts and decision-makers from HESE, during the exploratory individual interviews. The evaluation aspects, descriptors of performance, and reference levels were also created and, for that, a Workshop was conducted with the decision-makers (two healthcare professionals from HESE). After identifying the aspects, an analysis of interdependence between them was conducted, resulting in the criteria. For a better organization, the criteria were presented in a value tree, constructed through the M-MACBETH software. Regarding the performance of biomarkers in the various aspects, it was advised by the decision-makers to consult other professionals, namely from the Pathology and Medical oncology services from HESE, given the deeper knowledge of these professionals in the area. For that, two meetings with these two professionals were conducted.

After the structuring of the model, it was then possible to advance to the next stage: model building. During this stage, value scales were built, and weights were defined to the criteria. To do so, a Decision Conference was held with the decision-makers, in which the WisedOn software [70], a decision support system, was chosen to aid with the process.

As a result, a provisional multicriteria evaluation model is created. Thus, the next step was to test and validate the model with the decision-makers, which was also done during the Decision Conference. First, and using the value scales previously build, the biomarker options were evaluated, for each criterion, so that the performance of the options in each criterion could be converted into (1) a partial value score; and then (2) a global score, for the evaluation of the options, *i.e.*, add the partial value scores of the biomarkers across the criteria. After obtaining the overall scores for each biomarker, several sensitivity analysis were conducted, based on the results obtained.

Only after the phase of model testing and validation, the final multicriteria model is constructed, and it was possible to obtain the performance profile for each one of the options, as well as the ordered list of biomarkers, ranked by their respective scores. The most relevant biomarkers were presented to the decision-makers as the most promising biomarker options to be adopted in the near future by HESE to

Socio-technical approach using MACBETH

Technical component	Social component

Evidence synthesis

Rapid review for informing selection of biomarkers
+
Synthesis of stakeholders and expert's opinion results of a previous Web-delphi process for informing the selection of evaluation aspects

Model structuring

Definition of the decision problem at HESE

Biomarker selection

Definition of the evaluation criteria

Definition of the descriptors of performance, and reference levels

Assessment of the biomarkers' performance

Individual interviews with experts and decision-makers from HESE

Workshop Part I

Workshop Part II

Meetings with pathology and medical oncology experts

Model building

Construction of value scales

Definition of weights

Decision Conference Part I

Model testing and validation

Analysis of the biomarkers' overall value

Sensitivity analysis

Model validation

Decision Conference Part II

Final multicriteria evaluation model
For assistance in decision-making regarding the selection of emerging biomarkers to be adopted by HESE

Figure 4.1: Socio-technical approach that describes this thesis methodology. The left arrow indicates the flow of time, from top to bottom, culminating in the final multicriteria evaluation model.

handle HER2+ breast cancer patients.

Lastly, and as the final and last contact with HESE, it was asked, also during the Decision Conference, for the decision-makers to provide some feedback regarding the distinct stages of the socio-technical approach in which they participated, to gather general thoughts and opinions that arose

during the process, as a way to validate the approach.

To develop such methodology, it was important to establish a facilitation team, a group of people responsible to manage all the processes and activities associated with the MCDA methodology, and that would establish direct contact with the participants and the decision-makers during the individual interviews, the Workshop, and the Decision Conference (*i.e.*, in the social component of the process). The facilitator team for this methodology consisted of two persons.

Ahead, the MACBETH method will be explained in detail, to clarify some terms and to provide an easy understanding of the methodology. After the MACBETH subsection, the following sections will describe, thoroughly, the different steps associated with the methodology that was used, bearing in mind both the technical and social components. They are presented in four distinct sections: evidence synthesis, model structuring, model building, and model testing and validation, in agreement with the methodology phases (Figure 4.1).

4.1.1   The MACBETH multicriteria value measurement method

As already referred, the MACBETH method was used in this methodology.

MACBETH (Measuring Attractiveness by a Category-Based Evaluation Technique) is an MCDA technique, described as a "decision-aid approach for multicriteria value measurements" [50] that uses qualitative judgements to assess the difference in attractiveness to compare options in a set.

This approach is based on the additive value model, used to prioritize and select the distinct options available, and can be described through the following equation:

$$V(a) = \sum_{j=1}^{n} w_j v_j(a),$$
(4.1)

where $V(a)$ represents the overall value of alternative a, $v_j(a)$ translates the partial value of alternative a in terms of the criterion j and w_j is defined as the weight of the criterion j [50].

The additive model must also meet the following conditions (typically, if one considers the *neutral good* reference levels, explained ahead) [50]:

$$\begin{cases} v_j(good_j) = 100, \forall_j \\ v_j(neutral_j) = 0, \forall_j \\ V(good\ overall) = 100 \\ V(neutral\ overall) = 0 \end{cases}$$
(4.2)

$$\sum_{j=1}^{n} w_j = 1, w_j > 0 \quad \text{with} \quad j = 1, ..., n$$
(4.3)

Bellow, the distinct steps and concepts related to the MACBETH approach will be explained and will serve as a base for a better understanding of the next section (section 4.2), in which the implementation of the methodology will be performed.

Aspects

Once the problem is defined, the following step is to identify aspects that are relevant for the evaluation of the options. Each aspect consists of a Point of View (PV). The aspects typically emerge directly through the discussion with the actors, regarding what they consider to be relevant to the evaluation of the different options [71]. An aspect or PV, also referred to as concern, is any aspect that emerges during the discussion and is considered relevant to evaluate the options; is in the

consideration of at least one actors and lastly, has a well-defined and value meaning, that can be understood by everyone [72].

Criteria

A criterion, also referred to as Fundamental Point of View (FPV), corresponds to an individual PV or a set of PVs that are considered relevant for the evaluation of the distinct options, reflecting a fundamental value [71]. Each different FPV identified will correspond to an evaluation criterion in the multicriteria model. These criteria, or FPVs, must always have a specific set of characteristics. Each criterion must always be:

- **Intelligible**, allowing a clear understanding of the criterion/FPV;

- **Consensual**, translating concordance regarding the criterion/FPV between the actors in the decision-making process and also the facilitator(s);

- **Isolable**, to guarantee independence between the distinct criterion/FPV;

- **Operational**, to render the collection of information required for a reasonable analysis, considering both the time and effort available.

When considering a family of FPVs, other characteristics must be respected for it to be considered a "coherent family of criteria" (or FPVs) [71]. Thus, a family of FPV must respect the following desired properties [71]:

- **Complete** (or **exhaustive**), related with the fact that all the important and crucial aspects must be considered while maintaining a simple model at the same time;

- **Non-redundant**, guaranteeing that two or more criterion/FPV do not measure the same factor;

- **Concise**, contemplating only the hypothesis that is relevant to the model;

- **Decomposable**, allowing to be analysed separately;

- **Consensual**, translating concordance between the actors in the decision-making process and also the facilitator(s).

Once all the FPVs are identified, it is common to organize them and the areas of concern in an arborescent structure – typically called a value tree. This organizational structure allows for more intuitive visualization of the FPVs and their specification levels. Note that FPVs may not appear at the same level, since visualization of the PV might be desired [71].

Descriptors of Performance

Following the definition of the criteria/FPVs, it is necessary to associate, to each criterion, a descriptor of performance (also referred to as descriptor of impacts), for the criteria to be operational. Each descriptor is an ordered set of plausible performance levels and expresses, in a variety of ways, the criteria [73]. There is a great variety of descriptors of performance that can be classified according to three distinct dimensions, namely:

- **Quantitative**, **qualitative**, or **pictorial**: a quantitative descriptor relies on the use of numbers to create the performance levels; a qualitative one uses semantic expressions and numbers, while a pictorial one relies on visual representations;

- **Direct**, **indirect** or **constructed**: direct descriptors directly reflect the associated effects, since it is related with the criterion in a natural way; indirect ones indicate causes, as they are not good to translate the effects associated; constructed descriptors describe characteristics underlying the criterion;

- **Continuous** or **discrete**: continuous descriptors are represented by a continuous function, while discrete ones are represented by a finite set of levels.

Some descriptors are better than others, according to their characteristics. Generally speaking, a quantitative descriptor is always better to use than a qualitative one, if it is possible to use a quantitative one; the same applies to continuous descriptors, which are always preferred over discrete ones. If there is a direct descriptor, it should also be chosen [72].

Nevertheless, it is sometimes impossible to choose the best/ideal descriptor, as suggested, due to their complexity and impossibility to translate a criterion in a simple, direct way. That is why sometimes there is the need to construct descriptors of performance that are not, theoretically, the ideal ones (*i.e.*, quantitative, direct, and continuous descriptors of performance).

Reference Values

Besides the definition of the criteria and the descriptors of performance, it is also necessary to construct two reference levels, for each descriptor of performance. One of the most common methods to identify these "anchors" is to consider a *neutral* and a *good* level. A *neutral* level corresponds to a level of performance that is neither attractive nor unattractive: it is only acceptable; a *good* level corresponds to what is considered a good performance in the criterion [73]. Typically these two levels are associated with a score of 0 and 100 on the value scale (for the *neutral* and the *good* reference levels, respectively).

Despite the *neutral* and *good* intrinsic reference levels being recommended, one can also use other reference levels, such as the *worst* and *best* reference levels [50]. Nevertheless, it is advisable to use, when possible, the *neutral* and *good* reference levels instead of the *worst* and *best* ones. There are at least three different reasons to opt for *neutral* and *good* reference levels [71] [72]. First, the effort that is necessary to define the *neutral* and *good* reference levels contributes to improving the intelligibility/understanding of the FPV. Another reason is associated with the fact that, with the *neutral* and *good* reference levels, it is possible to classify the options with intrinsic value, *i.e.*, an option can be defined as a:

- *Very attractive option* – an option is, at least, as attractive as the fictitious option classified as *good* in all its FPVs;

- *Attractive option* – an option is, at least, as attractive as the fictitious option classified as *neutral* in all its FPVs;

- *Non-attractive option* – an option is less attractive than the fictitious option classified as *neutral* in all its FPV.

By using these reference levels, it is possible to explicitly determine the intrinsic value of each option, avoiding choosing an inadequate option, for the sole reason that it is the best one in a set of options (which would happen if the *worst* and *best* reference levels were implemented). Lastly, defining *neutral* and *good* reference levels allows not only the use of a valid weighting procedure, in the context of the additive model, but also avoids some of the pitfalls that exist in other classical weighting procedures [[74]].

After having defined the FPVs and their descriptors of performance, it is then possible to start constructing the model, using a decision support system, such as the M-MACBETH software.

Value Scales

The next step is to construct the value scales, after all FPVs and their descriptors are defined. This can be done by using the M-MACBETH software (note that in this thesis, it was used the WisedOn platform, but the same applies).

In the M-MACBETH software, for each criterion, there is a matrix of judgements associated. To fill in the matrix, the facilitator asks the decision-makers for differences in attractiveness between two distinct levels of the same criteria (*"What is the difference of attractiveness between these two levels of the criteria? And these two?"*, and so on, until all pairwise comparisons are performed). The scale used in the software is qualitative, *i.e.*, the decision-makers will use a set of semantic categories to classify the difference of attractiveness: *very weak*, *weak*, *moderate*, *strong*, *very strong*, or *extreme* [75]. The option of no difference (*null*) is also allowed in the software. While the matrix is being populated (only necessary to fill in the upper triangular part, without the necessity to fill in all the cells in the upper triangular part), the software checks its consistency each time a new judgement enters the matrix. If any inconsistency is detected, a suggestion for alteration is proposed (note that in the WisedOn software, there is no proposed suggestion; however, the software does not allow the facilitator to proceed with the protocol until a coherent judgement is finally inserted).

As mentioned above, for a set of *n* elements, it is not mandatory to fill all cells to populate the upper triangular matrix and perform all the paired comparisons ($n(n-1)/2$). In fact, the minimal number of judgements that is required is only $n-1$, as stated in [50]. Nevertheless, it is still advisable to perform additional judgements, for example, by filling the first two diagonals of the matrix, since extra judgements will allow a more reliable value scale [50].

Once the matrix is completed and is consistent, the qualitative judgements used to fill in the matrix are then used to generate a value scale for each criterion, by linear programming [75]. The scoring scale generated is numeric, so that the *good* performance level always corresponds to 100 and the *neutral* one always corresponds to 0, as previously mentioned. This way, there can also be negative values and values above 100, if these levels are below the *neutral* reference level or above the *good* reference level, respectively.

In some cases, and as already referred, it is not possible to assign a *neutral* and a *good* reference level. Hence, the *worst* and *best* reference levels are used.

After constructing the value scales, they must be adjusted by the decision-makers before the final approval (if necessary) and validated.

Weighting the criteria

In addition to the determination of the value scales for each criterion, it is also necessary to ascertain the relative weight of each criterion.

Calculating the weights is of extreme importance since they indicate the extent to which each criterion contributes to the final decision and are used to calculate the overall scoring of each one of the options, to reach the best decision. The weights must also be associated with the importance of the improvements, rather than just in terms of the importance of the criteria, as that would be considered "the most critical mistake" [50].

Therefore, to order the weights, the facilitator asks the decision-makers the following question: *"Suppose there is one option with neutral performances in all criteria. If its performance on one criterion could be improved to good, on which criterion would the improvement from neutral to good be the most important? And after that?"* [50], until there is no other criterion available. To answer this question, the decision-makers must reply using the same set of semantic categories of difference of attractiveness (*null, very weak, weak, moderate, strong, very strong, and extreme*) according to each criterion. This allows the filling of a weighting judgement matrix in the M-MACBETH software, using a similar procedure to the one used to generate the value scales. In case the decision-makers give an inconsistent judgement, the software does not allow it, and it suggests that the decision-makers provide another judgement. By linear programming, the software then computes the weights and the overall score of each option can be acquired, through the additive model, as will be explained next.

37

Global Score

After the construction of the model, it is now possible to test and evaluated it, followed by a validation of the model. This is done by inserting the different options and the options' performance in the software.

Subsequently, the weighting coefficients, k_j for each criterion, which define the weighting that each one will have for the final value, are computed. They allow each partial value unit, v_j, to be converted to a global value unit, V. Thus, one can mathematically determine the global value for each alternative, a, through the additive model described in equation 4.1, while still respecting the conditions referred to in equation 4.2 and equation 4.3.

Thus, and after all the necessary information is collected and filled in the model, the quantitative evaluation model is constructed, which generates a numerical scale consistent with the decisions taken along the process. The overall score of each option is calculated, considering the sum of the partial weights of each option in each criterion, as shown above. The option with the highest global score, or the top options with the higher global scores, will be the most attractive for the decision-makers.

Sensitivity Analysis

In projects of this nature, it is important to keep in mind that the judgments given by the decision-makers are given in a qualitative and somewhat uncertain way. Thus, given the subjectivity of the collected data, the obtained results can have an associated margin of error that varies from criterion to criterion.

Therefore, it is indispensable to perform a sensitivity analysis to weight variations within the ranges allowed by the judgments made. Hence, one can define the sensitivity analysis on the weight of a criterion as a tool that allows the evaluation of the change in the final results of the constructed model [76]. Sensitivity analysis for each criterion should be conducted, to determine if relatively small changes on the criterion weight could affect and modify the results obtained.

By direct observation of the example in Figure 4.2, it is possible to see distinct straight lines. The lines depicted on the graphic represent the overall variation of a given option when the weight of the criterion varies on a percentage scale (from 0% to 100%) while the vertical red line indicates the current weight value for the criterion chosen in the analysis, after applying the additive model.
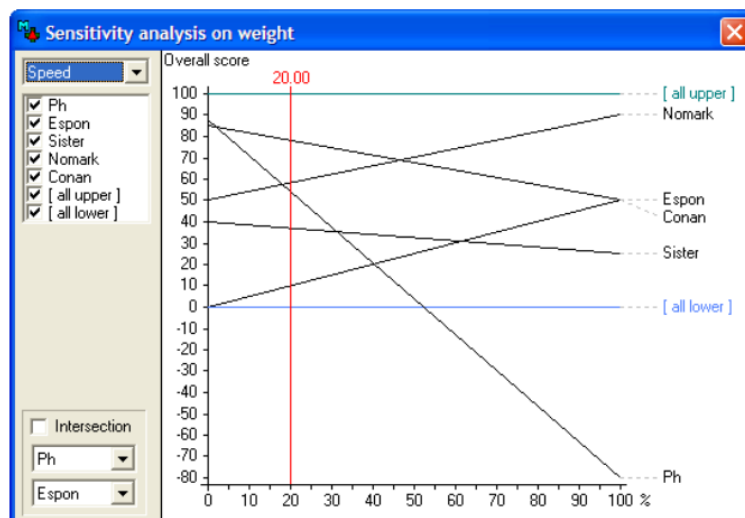
Figure 4.2: Example of a sensitivity analysis graphic. Note that, in this example, the reference levels chosen were the *worst* (all lower) and *best* (all upper). [source: M-MACBETH software [76]]

Robustness Analysis

After performing sensitivity analysis, which deals with the possible variation of the weight of the different criteria, one can also consider the uncertainty related to any decision-making process: it is a result of a human personal and biased opinion, which can never be perfectly expressed to the analyst. A robust commitment will hold dominance and global preference between options under varying amounts of information.

Robustness analysis can be applied when two requisites are met [76]:

1. Uncertainty (which may be originated from incomplete, imprecise, or uncertain information) is a factor that thwarts the confidence of the decision-maker: this ensures that uncertainty must be taken into account. In this case, uncertainty is surely present, not only because there cannot be a perfect communication between the decision-maker and the facilitator, but the decision-makers themselves are sometimes indecisive about where to classify certain options in the scoring scale of the different criteria;

2. Commitments made in the initial stages of the decision process (for example, choosing a score of performance of an option in certain criteria) do not necessarily define completely the future state of the system, that is, the final decision. In other words, there must be other criteria that also influence the final result.

It is possible to explore the robustness of the decision in the M-MACBETH software through its robustness analysis function, which shows the results in a matrix, where the different options can be successively compared (although this analysis was not performed in this thesis). The matrix positions are given by their (i, j) coordinates. The symbols below (also depicted in Figure 4.3) aim to explain the connections between them [76]:

$\triangle$ This symbol represents total dominance, *i.e.*, option i is, at least, equally attractive to option j in all criteria and more attractive in at least one of these;

$+$ This symbol represents additive dominance, *i.e.*, option i is always more attractive globally than option j considering the model applied (because of that, additive dominance is more sensitive to uncertainty than total dominance);

$=$ This symbol translates no dominance between options, *i.e.*, both options are equally attractive in every criterion;

$?$ This symbol is present when there is insufficient information, *i.e.*, the available information is not sufficient to determine dominance between the options i and j.

Figure 4.3: Example of a robustness analysis matrix. [source: M-MACBETH software]

4.2 Implementation of the Proposed Methodology

As already mentioned, in this section, a detailed description of the various phases of the proposed methodology will be presented, based on the information provided in the previous subsection (subsection 4.1.1).

The section is divided into the four phases required for the development of this methodology: evidence synthesis, model structuring, model building, and model testing and validation.

During all these steps, the decision-makers Dr. Hugo Quintino and Dr. Jorge Caravana were present, although in the methodology there were also inputs provided by other healthcare professionals from HESE, that will be detailed ahead. Nevertheless, only the two decision-makers were responsible for making the decisions, and every input and opinion given by the other participants was always provided to the decision-makers for their approval.

4.2.1 Evidence synthesis

During the first stage of the methodology, the evidence synthesis stage, two actions were conducted in parallel:

- a rapid review, to inform the selection of emerging biomarkers for HER2+ breast cancer;
- synthesis of stakeholders and experts opinions results based on a previous Web-Delphi process, to inform the selection of evaluation aspects.

Ahead, a more detailed explanation regarding both will be presented.

Rapid review for biomarker selection

According to the literature and to what was reported by the medical doctors at HESE, there is an increasing need for new biomarkers to be used in the clinical context, for this subtype of breast cancer, not only to provide better diagnostics and prognostics but also to lead to better and personalized treatments.

Due to interest demonstrated by the healthcare professionals at HESE, a rapid review regarding emerging and promising biomarkers for HER2+ breast cancer was performed. This rapid review used as a base a previous literature review concerning emerging biomarkers for breast cancer, with no particular focus on the HER2+ subtype, since, at that time, the healthcare professionals at HESE wanted to encompass emerging biomarkers for all breast cancer subtypes.

Therefore, the initial search conducted had an aim for secondary research studies (*i.e.*, reviews and systematic reviews). A search of the PubMed [20] and ScienceDirect [77] databases was conducted, using the keywords *breast cancer*, *biomarkers*, *diagnostic*, *prognostic*, and *drug response*. It was decided that only these terms would be used to maximize the number of retrieved results. Search simulations were conducted with different terms and specific searches until it was ensured that the results obtained would be pertinent and in line with the main goals of the search.

This search had the focus for biomarkers that are currently still not validated for therapy decision-making by the majority of guidelines (and thus, are not used in the clinical context), and that are present in recent literature (ranging from the years 2014 to 2020) so that it is possible to deal with emergent biomarkers. Despite being still relatively new, they hold great promise in becoming potential tools to optimize the current diagnostic, prognostic, and treatment in breast cancer.

By searching with the aforementioned keywords, a total of 12.441 articles was retrieved, including the searches in the PubMed and ScienceDirect databases and 3 other articles (these 3 systematic reviews were obtained through the Web search engine Google and a scientific journal). To initially filter these

results, the majority of them were excluded, since they were articles that did not correspond to reviews or systematic reviews; the publication years were out of the interval that was initially set; the articles were not in English or the articles did not present full-text availability. After this process, 2.134 articles were obtained. These articles were once again filtered, this time by title and abstract review. All of the articles considered off-topic were immediately rejected. Cases of duplicates were also found and removed. This left 116 articles to undergo the last scrutiny, stricter, in which any reviews (systematic or non-systematic) were removed if considered with low quality or not relevant. Finally and after this process, 29 articles were considered and fully analysed (Figure 4.4).
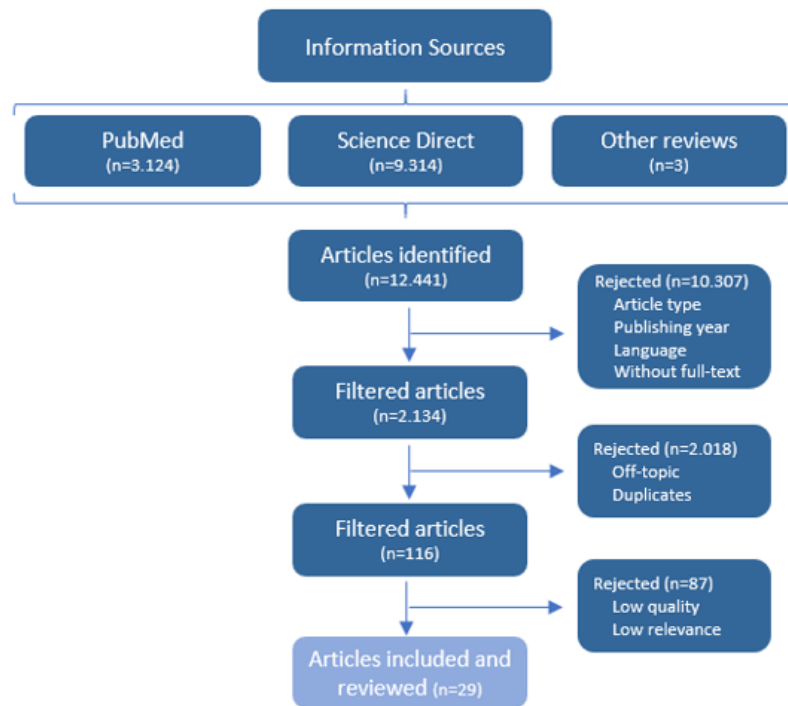
Figure 4.4: Literature review search flow, resulting in 29 articles to be fully analysed. The resulting pool of articles served as a base for the rapid review with the focus on emerging biomarkers for HER2+ breast cancer.

It is important to stress once again that, in this phase, it was still not defined that the biomarkers to be analysed would only consist of the ones associated with HER2+ breast cancer. Therefore, and when the focus was defined, the same pool of the resulting 29 articles was once again analysed, this time with a well-defined focus for the HER2+ breast cancer subtype. The results obtained specifically for this breast cancer subtype were then presented to the healthcare professionals from HESE for approval. The biomarkers were organized in four distinct categories: HER2 & Human Epidermal Growth Factor Receptor 3 (HER3), Gene expression, DNA mutations, and Immune microenvironment, presented in Table 4.1. This category organization did not interfere with the results, since it was only a way to present the biomarkers to the healthcare professionals in a more arranged form.

Table 4.1: List of the emerging biomarkers' options to be used for HER2+ breast cancer, grouped in distinct categories, after approval from the healthcare professionals at HESE.

Category	Biomarker	Reference
HER2 & HER3	HER2 levels	[64], [78]
	HER2 mutations	[66]
	HER2 heterogeneity	[64]
	HER3	[64]
Gene expression	Intrinsic subtype (PAM50)	[64], [79]
	Phosphatase and tensin homolog (PTEN)	[80], [81]
DNA mutations	PIK3CA mutations and PI3K pathway inhibitors	[64], [65], [66], [78], [80], [81], [82], [83]
Immune micro-environment	TILs	[64], [66], [84], [85]
	PD-L1	[64], [86], [87]
	FcγRs	[64], [66]
	Liquid biopsy (ct-DNA, ct-miRNA, CTC)	[66], [88]

Synthesis of stakeholders and experts opinion (using a previous Web-Delphi process)

Also during this first stage, aspects from a Web-Delphi process, regarding a MEDI-VALUE study ("Identificação de aspetos relevantes para a avaliação de dispositivos médicos - o caso particular dos Testes *in vitro* baseados em biomarcadores e Dispositivos médicos implantáveis") were used and integrated in this thesis.

The Web-Delphi was conducted to gather the views of a wide and diverse group of stakeholders to assess which aspects are potentially relevant for the evaluation of medical devices, in the specific context of implantable medical devices and *in vitro* tests based on biomarkers. The Web-Delphi was elapsed between March and May 2020 and counted with the participation of 167 and 134 participants, in the 1st and 2nd rounds, respectively. Participants were also distributed among four distinct stakeholder groups: (1) health professionals (doctors, nurses, pharmacists, senior technicians); (2) buyers, policymakers, and academics; (3) industry; and (4) patients and citizens.

The aspects considered in the Web-Delphi were used as the base and starting point for this thesis, *i.e.*, by using the aspects mentioned in the Web-Delphi, it was possible to collect information from the health professionals at HESE regarding which of them they considered important for the evaluation of *in vitro* tests based on biomarkers for HER2+ breast cancer.

4.2.2 Model structuring

The following stage of the methodology consists of the structuring of the model. It was during this phase that the criteria were defined (from the initial aspects gathered in a Web-Delphi from a MEDI-VALUE study), and descriptors of performance were associated with each aspect.

In parallel, and to understand which aspects were considered relevant for the evaluation of *in vitro* tests based on biomarkers for HER2+ breast cancer, individual interviews with eight healthcare professionals from different services of HESE were carried out, based on the results of the Web-Delphi process.

After that, a Workshop was performed, to define the criteria and the descriptors of performance associated with each criterion. This was conducted with a more narrow subgroup of the eight healthcare professionals: two professionals, that were considered the decision-makers and were always present, from the initial to the final phases.

Thus, each of the activities done in this phase will be explored in detail, in the following subsections.

Individual interviews with experts and decision-makers

As already mentioned, a Web-Delphi, inserted in the MEDI-VALUE project was used. To assess the relevant aspects for the specific case of HER2+ breast cancer, individual interviews were conducted, in an informal manner, with a total of eight healthcare professionals from HESE. The choice of the participants was of extreme relevance since they will directly translate the validity of the results obtained afterwards. It was also important to choose healthcare professionals from the hospital's distinct services (namely Pathology, General surgery, Medical oncology, and Hospital administration) to gather varied insights and opinions (Figure 4.5).
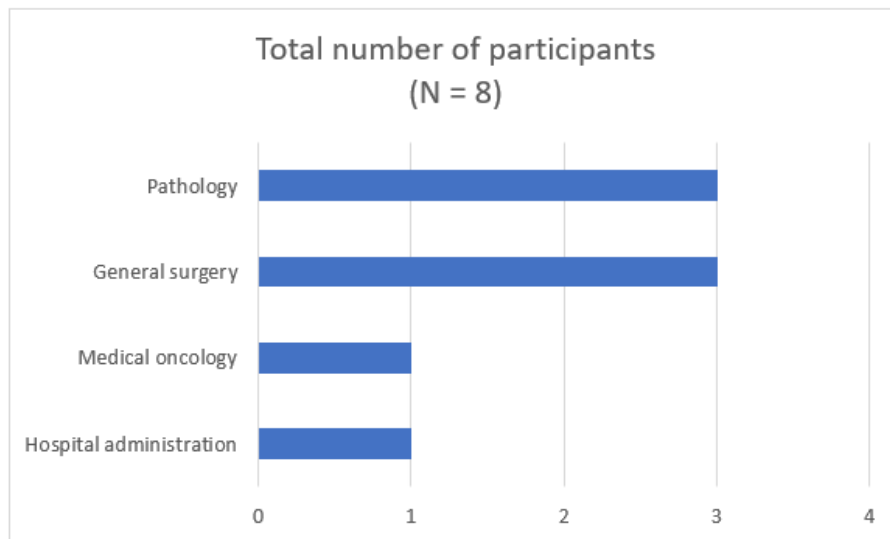
Figure 4.5: Distribution of the eight participants (medical doctors at HESE) in the individual exploratory interviews, by the different hospital services.

The exploratory individual interviews took place between August 13th and September 14th, 2020, via phone call, since it was the easiest and most intuitive way to communicate with the clinicians, in the current context of the COVID-19 pandemic. The interviews were semi-structured. Structured since they were based on the classification of different aspects regarding its relevance in *in vitro* tests based on biomarkers for HER2+ breast cancer disease. However, there was still space for a less structured part, since participants were free to ask questions to clarify any doubts that might arise, add comments or any other information, in case they thought it was necessary or relevant.

The classification method used to classify the distinct aspects, in the Web-Delphi, was also re-used for this thesis, allowing, each aspect could be classified based on an already established answer scale (relevance scale). Thus, the aspects could be classified by the healthcare professionals as:

- **Critical**: meaning that the aspect beyond fundamental can, by itself, preclude assessing if the medical device has added value given its alternative;

- **Fundamental**: the aspect must, undoubtedly, be part of the basis of evaluation of the medical device to assess if it has added value given its alternative;

- **Complementary**: the aspect is not fundamental but, still, it can add something to the value of the medical device given its alternative;

- **Irrelevant**: the aspect must not be part of the basis of evaluation of the medical device; it is inapplicable or irrelative to assess if the medical device has added value given its alternative;

- **Do not know/ Do not want to answer**: in case the participant was not comfortable in classifying the aspect.

Prior to the interviews, it was sent to each participant two documents, via email: the list of the emerging biomarkers for HER2+ breast cancer, resulting from the rapid review (an initially larger version of Table 4.1, which was later reduced after their input) and another one with the aspects obtained from the Web-Delphi and their respective description (Table 4.2).

Table 4.2: The 34 aspects associated with the evaluation of *in vitro* tests based on biomarkers and its descriptions (obtained from the results of a Web-Dephi, prior to the development of this thesis).

Aspect	Aspect Description
Specific characteristics of the medical device	Relevant characteristics such as: lifetime of the medical device, contact with the patient's body (direct: invasive and non-invasive; indirect), length of hospital stay; limitations compared to alternatives
Technical performance of the medical device	Behavior of a medical device or response of the patient(s) to the medical device regarding its intended purpose and as indicated by the manufacturer, when properly applied to the appropriate patient(s)
Regulatory status of the medical device	Current regulatory status of the medical device and adherence to market approval requirements
Sensitivity and specificity	Ability of the medical device to correctly identify positives and negatives in testing a population
Ease of use for healthcare professionals	Extent to which the medical device is comfortable and easy to be used by the healthcare professional [*e.g.*, depth perception, eye-hand coordination or hours required to perform the examinations (compared to the expected time)]. The procedures for using the medical device are defined and clear. Extent to which the result of the medical device is easy to understand and interpret
Time between procedures and results	Extent to which medical device results are instantly available
Need for health professionals training	Need for training of health professionals to ensure specific knowledge necessary for the use of the medical device
Health professional's learning curve	Improvements in the use of the medical device by the healthcare professional (*e.g.*, evolution of speed and accuracy of use)
Exposure of the health professional to physical or chemical agents	Exposure (internal or external) of the health professional to specific physical and chemical agents, metabolites, or reaction products
Workload for healthcare professionals	Extent to which the medical device can be implemented without drastic changes in the current way of working and in the workload of healthcare professionals
Patient's Comfort	Extent to which the use of the medical device is expected to improve the service offered to the patient (includes patient satisfaction)
Connectivity	Ability to transmit information between different devices and decision support, including security solutions in the transmission of biosignals
Efficacy and/or clinical effectiveness	Measure of the beneficial effect of a technology, *i.e.*, the medical device's ability to produce a desired (beneficial) change in the signs, symptoms and evolution of the target health condition
Risk analysis	Considers risk factors related to the potential number of patients at risk of harm: the likelihood of a medical device having problems, the likelihood of a patient being harmed, and the total number of patients exposed
Adverse events for the patient	Seriousness and frequency of adverse events and their impacts
Quality of the available scientific evidence	Robustness of the sources of evidence and quality of the evidence, with regard to the target population and the intended clinical course, in a complete, consistent and relevant way, with the problems properly reported (risk of bias, imprecision, inconsistency, indirect evidence and publication bias)

Target population	Number of people affected by the relevant condition or subgroup(s) in a specific time period
Disease impact	Impact of health condition on the disease burden for the patient and society, the incidence and prevalence and the natural history of the disease without treatment
Patient-reported outcomes	Patient's perspective on a disease/treatment that may not be captured by a clinical measure, but of importance to the patient and their adherence to treatment or use of the medical device
Quality of life for the patient	Expected impact on the patient's quality of life
Space for innovation in the organization	Ability to accommodate the medical device
Clinical guidelines	Extent to which the medical device is recommended in national clinical guidelines
Financing	Extent to which the medical device is reimbursed; additional payments for the medical device (coverage)
Public health concern	Impact of the use of the medical device on public health and society
Impact on health system budget	Impact of medical device adoption on health care budget
Equity	Justice in the allocation of resources, treatments or results between different individuals or groups
Market competitiveness	Competition in the market, in terms of the number of competing products
Medical or technical complications for the patient	Expected occurrence and severity of medical and technical complications related to the procedure (for example, with surgery to implant the medical device)
Agreement of key actors in the adoption of the medical device	Agreement between key actors (groups of stakeholders) or individuals in the health context, that is, the extent to which the adoption of the medical device is aligned with common goals
Environmental impact on the use of the medical device	Extent to which the production and use or implementation of the medical device causes environmental damage
Efficiency	Potential cost savings with the adoption of the medical device, considering the affordability of the medical device and the opportunity cost (lost alternatives)
Health system capacity	Adjustment of the medical device to the existing infrastructures and qualifications in the health system
Medical device cost (including complementary equipment)	Costs with the medical device (*e.g.* disposable items) and other equipment inherent in the use of the medical device, maintenance (sterilization, licensing or updates) and training
Cost of the procedure without the cost of the medical device	Costs arising from the number of human resources required, time of use and infrastructure, for example, operating room (without considering the cost of the medical device)

Since the order and way of asking questions can interfere with the type of answer the interviewee provides, the interview process was structured the following way:

1. In one of the documents sent to you via email, you can find aspects that can be considered for the evaluation of *in vitro* tests based on biomarkers. For the specific context in which we are inserted, *i.e.*, *in vitro* tests based on biomarkers for HER2+ breast cancer disease, which ones do you:

   - Consider *relevant* and *irrelevant*?
   - Regarding the ones you considered *relevant*, can you classify them as *Critical*, *Fundamental*, and *Complementary*?

2. Do you consider that the presented aspects are of easy understanding, regarding your knowledge on the area? If not, what do you think could be improved?

3. Within the mentioned aspects, do you consider that there is a missing aspect that was not contemplated, but should be included?

4. Do you approve the list sent to you, with the potential biomarkers that can be used for HER2+ breast cancer disease?

The conducted interviews were relatively short in time (between 15 to 30 minutes) and allowed the participants to be comfortable to speak regarding any issue that they might consider relevant. The order of the questions presented above was respected, leading to the final question associated with the approval of the emerging biomarkers list to be used in HER2+ breast cancer disease (extended version of Table 4.1). After this last question, the interview was concluded.

Workshop

After the individual interviews, the next contact held with HESE was through a Workshop.

This subsection provides a deeper focus on the Workshop developed with two healthcare professionals, from HESE - Dr. Hugo Quintino and Dr. Jorge Caravana - session in which the model was structured. They were also participants during the exploratory individual interviews. Due to the participants' time constraints, the Workshop was split into two parts, conducted in two distinct days: Workshop Part I and Workshop Part II, detailed next.

Workshop Part I

Once the first phase of exploratory individual interviews ended, the succeeding step consisted of a more detailed exploration of the results obtained during the interview sessions that consisted of the Workshop. Workshop Part I was held on November 19th, 2020 via *Zoom*, with the two healthcare professionals - Dr. Jorge Caravana from the General surgery service and Dr. Hugo Quintino, from the Hospital administration, from HESE. The Workshop Part I had a total duration of approximately 1 hour and 30 minutes.

During this phase, a reflection with the two professionals was performed, and a focus on the structure of the model was done, in this case, regarding which aspects are considered relevant for the evaluation of *in vitro* tests based on biomarkers for HER2+ breast cancer, for future application in the hospital context, at HESE.

As a starting point, at the beginning of the Workshop, there was a reflection regarding the information presented and the results obtained during the previous stage of the methodology, namely, the results of the Web-Delphi process from the MEDI-VALUE project, and the results of the individual interviews.

Distinct activities were planned for Workshop Part I. However, due to time constraints, it was only possible to resolve the following concerns:

1. Clarification of the focus of evaluation of the HESE case study;

2. Definition of the aspects to consider when evaluating *in vitro* tests based on biomarkers for HER2+ breast cancer. According to the results of the individual interviews:

   • Choice, for the specific context of HESE, of which aspects should be considered.

Regarding the first activity, it was necessary to identify the focus of the evaluation of biomarkers. After the rapid review process, the group of biomarkers suggested presents 3 distinct functions: biomarkers for diagnosis, prognosis, or prediction of drug/therapy response. Each biomarker suggested can have

46

one, two, or all three of these functions. Therefore, participants were asked to clarify what the focus of interest was for this case study: whether all biomarkers collected in the literature or only a few focusing on some of the function(s) were going to be considered.

When considering the second activity, the healthcare professionals were asked to reflect on which aspects should be considered relevant to assess the set of emerging biomarkers previously selected, taking into account their perspectives and the outcome of the individual interviews, as well as the fact that there may not be enough evidence to analyze some aspects at this stage of adoption decision (it is relevant to stress that, during the individual interviews, the biomarkers' group for HER2+ breast cancer was not always taken into account by the participants, due to lack of particular knowledge for the HER2+ breast cancer subtype, in some cases. Instead, a more broad analysis regarding biomarkers for general breast cancer was done by some of the participants). During the classification of the different aspects, the decision-makers used the same relevance scale as the one described in section 4.2.2.

By doing this reflection with the healthcare professionals, it was possible to narrow down the relevant aspects from the initial list of the 34 aspects (Table 4.2).

Due to the time constraints and the availability of the health professionals, as already mentioned, it was not possible to explore in all the activities that were planned (including an in-depth analysis of overlaps and/or redundancies between aspects and the performance of the distinct biomarkers in the identified aspects).

Nevertheless, it was explained what was intended with these activities and it was agreed that material would be prepared after the Workshop Part I to be discussed with the participants, in the next session - Workshop Part II.

Workshop Part II

The second part of the Workshop consisted of the completion of the suggested activities from Workshop Part I. To conclude structuring the multicriteria model. This session took place on February 4th, 2021, with the same two healthcare professionals. Just like Workshop Part I, this session lasted approximately 1 hour and 30 minutes and was also held via *Zoom*.

The proposed activities for the Workshop Part II were the following:

1. Definition of the aspects to consider when evaluating *in vitro* tests based on biomarkers for HER2+ breast cancer. According to the results of the individual interviews:

   - Analysis of possible overlaps or redundancies between the distinct aspects;

2. Analysis and validation of the proposed descriptors for the different aspects;

3. Exploration of the performance of biomarkers in the identified aspects.

Before starting with the activities proposed for Workshop Part II, a review of the activities carried out in the previous Workshop Part I was done, to contextualize, as best as possible, the participants for this session.

As mentioned, the first activity of Workshop Part II was to identify potential overlaps, redundancies, and/or inter-dependencies between the different aspects. For this, a value tree was presented, constructed through the M-MACBETH software, in which the relevant aspects, resulting from Workshop Part I, were organized in different dimensions of evaluation. It was also mentioned during this exercise that the costs associated with the biomarker were only going to be considered afterwards, following a logic of *Value for Money*. At this stage, the decision-makers were asked to analyze the value tree and, if necessary, to suggest changes in its organization. They referred that the evaluation dimensions

considered and the aspect organization made sense, which allowed proceeding to the validation/discussion of some characteristics of the value in greater detail.

Then, the descriptors (or performance scales) built for each aspect were presented to the decision-makers. With this exercise, it was intended that the decision-makers gave their opinion, based on their experience in the area, regarding the performance scales previously built. Some of these scales did not undergo any changes and remained the same as those initially proposed. However, in the large majority, proposals for improvement and alteration were mentioned, including grouping of two aspects, due to redundancies between them; the removal of an aspect; and consideration of one aspect only at a later stage, following a perspective of *Value for Money* (in cases where the aspects were associated with monetary costs). Some of these performance scales were also rejected by the decision-makers. However, they had some doubts and reported not having enough knowledge to propose changes to the rejected performance scales. Thus, they both suggested that contact with professionals in the area of Pathological Anatomy at HESE should be done, for approval of certain performance scales or proposal of new ones, given the deeper knowledge of these recommended professionals, for this particular concern.

Lastly, and regarding the exploration of the performance of biomarkers in the various aspects: in some cases, the decision-makers stated that they did not have enough knowledge and information to associate the different biomarkers to a level, in some aspects (namely in those aspects related with the healthcare professional that handles the biomarkers, and technical characteristics of the biomarkers). This only reinforced the need to contact healthcare professionals from Pathology and Oncology at HESE, which allowed the reconfiguration of some of the performance scales as well as the biomarkers' performance in some aspects.

After the Workshop Part II and the conversation held with the responsible for the service of Pathology, Dr. Carlos Quintana, and Medical Oncology, Dr. Rui Dinis, it was necessary to perform changes in the proposed value tree presented at the beginning of the session.

4.2.3 Model building

The following stage of the methodology was to construct the model, *i.e.*, build the value scales and weight the criteria. This was possible to execute since the aspects considered relevant were already selected and can be referred to as criteria, at this point of the methodology, since all redundancies between them were already eliminated, during the Workshop Part II.

These activities were done by performing a Decision Conference with the same two decision-makers, which was designated as Decision Conference Part I.

Decision Conference Part I

The first part of the Decision Conference was carried out on March 29th, 2021, via *Zoom*. It had a total duration of approximately 2 hours.

As already mentioned, the two main activities were performed in this session were:

1. Construction of value scales for each criterion;
2. Assignment of a weight to each of the different aspects.

To carry out these activities, the WisedOn decision support system was used [70]. WisedOn is an online system that implements the Advanced Value Framework, a methodological framework for assessing the value of new medical technologies using the MCDA approach MACBETH [70].

In this thesis, the WisedOn software was preferred over the M-MACBETH software, since it provides a more user-friendly experience, with well-organized steps, clear, simple, and easy to explain to the decision-makers. After all, one must always bear in mind that the contacts with HESE professionals

(both the decision-makers and the other experts) were performed remotely, via phone call (during the individual interviews), and by *Zoom* (in all other sessions). Thus, it was important to have intuitive tools so that the process could be as streamlined as possible, since it was not possible to have direct, *in loco* contact with the participants.

Nonetheless and as already mentioned, the WisedOn software is based on the MACBETH approach and the steps executed in WisedOn are extremely identical to the ones carried out on the M-MACBETH software. Thus, the steps and concepts explored in Subsection 4.1.1 are also maintained while using WisedOn (except for the robustness analysis section, not available on the WisedOn software, to date). Regarding the first activity - constructing the value scales for each criterion - the decision-makers were asked to classify the difference in value/attractiveness between levels of the same criterion, by performing pair-wise comparisons. For example, the questions asked for a generic criterion C with three generic levels $l1$, $l2$, and $l3$ were the following:

- "For criterion C, how do you classify the difference in value between level $l1$ and level $l3$?"

- "For criterion C, how do you classify the difference in value between level $l2$ and level $l3$?"

- "For criterion C, how do you classify the difference in value between level $l1$ and level $l2$?"

The decision-makers had to reply accordingly to the MACBETH semantic scale *(null, very weak, weak, moderate, strong, very strong*, and *extreme*). When the decision-makers had doubts between two levels of the semantic scale, some hesitation was allowed, and the answer could be a combination of consecutive semantic levels (*e.g.*, a judgement of *moderate-strong*, in case the decision-makers were hesitant in classifying the difference in value as *moderate* or *strong*). The same questioning protocol was asked for all the other criteria, for each pair of levels, to build the matrix of judgements that would allow the creation of the value scales. Despite the current WisedOn default questioning protocol only asks for judgements associated with the filling of the last column of the matrix followed by the first diagonal, it was asked to the decision-maker to also fill the rest of the matrix so that all the cells had a judgement associated. So, after the filling of the first diagonal, the second diagonal was filled, followed by the third, fourth (if this was the case), and so on, until there was not any cell left.

After that, the generated scales were then presented to the decision-makers, so that they would modify the judgements given, in case they were not satisfied with the value scales or approve the, in case they agreed with the obtained scales.

The second and last objective of Decision Conference Part I was to assign a weight to each criterion. For that, and also using the WisedOn software, the swings for each criterion between its *good* and *neutral* levels were ranked in comparison with the rest of the criteria. Thus, it was asked to the decision-makers to order the criteria based on its *neutral*-to-*good* swing (from the most important swing to the least important one). For that, the protocol question was asked: "*Which one of these criteria swings do you consider the most important? And after that?*", until there were no criteria left to order.

Subsequently, the decision-makers were asked to assign a judgement to each of the *neutral*-to-*good* swings (that were previously ordered, Figure 4.6), and swing pairs (Figure 4.7), using qualitative judgements, that rely, once again, on the MACBETH semantic scale. The questioning protocol was the following, for both individual swings and swing pairs, respectively:

- For an individual swing (for a generic criterion C): "*How important to you is this improvement, regarding criterion C?*"
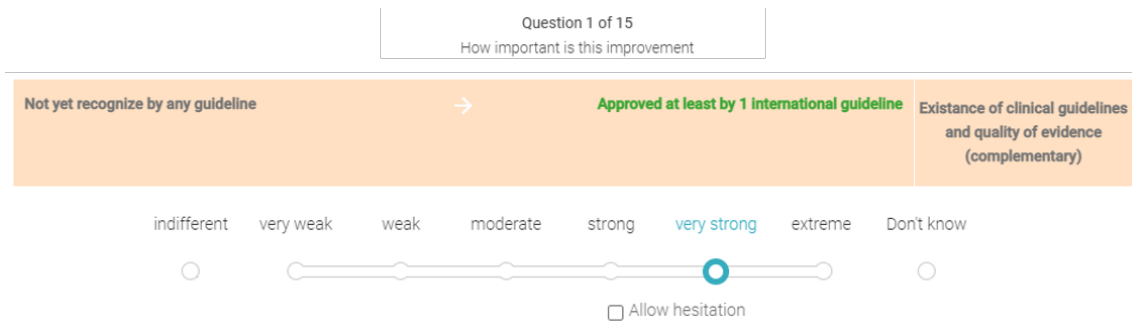
Figure 4.6: Example of a questioning protocol for calculating the weights of the criteria, in this case comparing how important would be an option with a good level in the aspect *Existence of clinical guidelines and quality of evidence*, and a neutral level in remaining aspects, compared to another option with all aspects at the neutral level. [source: WisedOn platform]

- For a paired comparison (for two generic criteria C_1 and C_2): *"How much more important is the improvement at the top, regarding criterion C_1, compared to the improvement at the bottom, regarding criterion C_2?"*
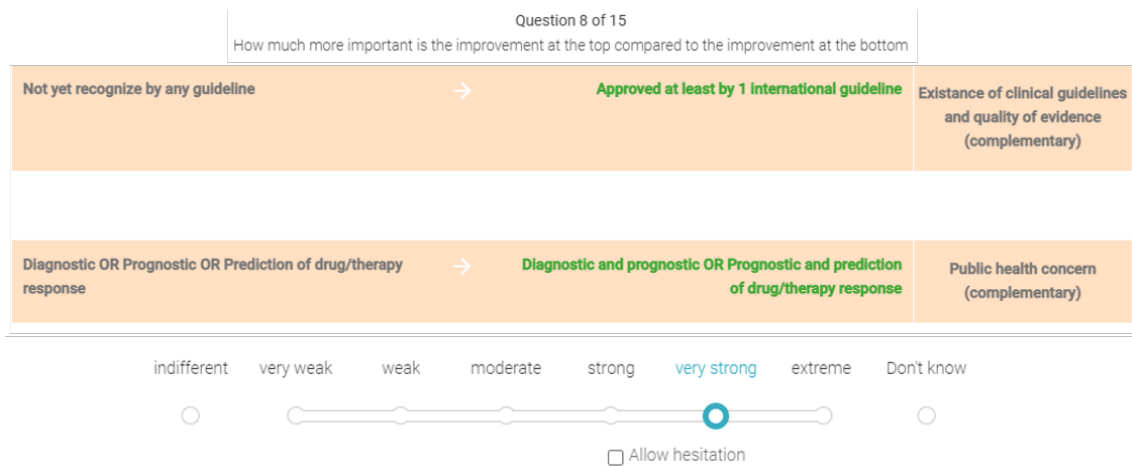
Figure 4.7: Example of a questioning protocol for calculating the weights of the criteria, in this case comparing how important would be an option with a good level in the aspect *Existence of clinical guidelines and quality of evidence*, and a neutral level in other aspects, compared to another option with a good level in the aspect *Interest for public health* and a neutral level in the others aspects. [source: WisedOn platform]

As a result, it was possible to fill in the weighting judgement matrix and the relative weights of each criterion were obtained. Note that, by using the WisedOn software, the default questioning protocol is similar to the one used in the filling of the matrices for the value scales. It starts with the judgements of the individual swings, which correspond to the filling of the last column of the weighting judgement matrix; followed by the filling of the matrix's diagonal by order: the first (judgements between two consecutive criteria), the second matrix diagonal, and so on, until all the judgements are inserted in the matrix.

Once again, it was also asked to the decision-makers to approve, if they were complacent, or to make some changes in the judgements, in case they did not agree with the generated weights.

It is important to note that, both during the value scales building and the criteria weights assigning, the decision-makers would sometimes provide an inconsistent judgement, *i.e.*, sometimes the decision-makers would not respect the conditions required to build the matrices of judgement and, as a

consequence, the judgement provided was not compatible with the others [50]. During these cases, the WisedOn software (just like what occurs in the M-MACBETH software) alarms the user that the judgment is not possible and must be changed to be in concordance with the other judgements already inserted in the matrix. Despite not being frequent, some of these inconsistent judgements occurred and changes were consequently made, with no difficulty on the part of the decision-makers.

4.2.4 Model testing and validation

After the construction of the model, it was finally possible to proceed to the last stage: model testing and validation. In order to test and validate the model, a final session with the decision-makers was conducted - the Decision Conference Part II, explained ahead.

Decision Conference Part II

The last direct contact with the decision-makers from HESE occurred during the Decision Conference Part II, a session that took place on 26th April 2021, with the same two decision-makers. It was also held via *Zoom* and it had a total duration of approximately 1 hour and 30 minutes.

This session served as a conclusion of the methodology that was conducted, and 3 activities were proposed and executed:

1. Analysis, comparison, and discussion of biomarker scores;

2. Sensitivity analysis;

3. Feedback/comments from the decision-makers regarding the methodology used.

Regarding the first activity, the results obtained from the Decision Conference Part I were presented to the decision-makers and discussed, namely the biomarkers scores and ranking, as well as comparisons between pairs of biomarkers regarding their performance in the different criteria. For the latter, the WisedOn software was used for pairwise profile analysis.

Next, sensitivity analyses were performed, also using the WisedOn software. Sensitivity analyses were executed for all criteria and in some cases, for some criteria, only a few biomarkers were chosen for a more detailed and focused analysis.

Lastly, it was asked for the decision-makers to provide some feedback regarding the methodology used and to add some other comments, opinions, and thoughts that they would like to share. The questions were asked to gather their feedback concerning both the structuring phase of the methodology, as well as the model building phase.

The results of all the phases described in this chapter will be presented in the subsequent chapter: Chapter 5.

Chapter 5

Results

In this chapter, the results obtained through the proposed methodology, described in the previous chapter, will be presented. For a better understanding, the following sections are organized as the sections presented in Chapter 4, *i.e.*, in the four distinct phases of the methodology: evidence synthesis, model structuring, model building, and model testing and validation.

5.1   Evidence synthesis

5.1.1   Rapid review for biomarker selection

As already described in Chapter 4, a rapid review was conducted to select the emerging biomarkers for HER2+ breast cancer that would be evaluated. In Table 5.1, the final list of the distinct biomarkers found through evidence synthesis, after the approval of the healthcare professionals at HESE, is presented. During this first stage, the biomarkers were grouped into four distinct categories, according to their characteristics (HER2 & HER3, Gene expression, DNA mutations and Immune micro-environment). As previously stated, the categories served only as an organization of the biomarkers, for a better presentation to the healthcare professionals, and did not interfere with the consequent results obtained.

Table 5.1: List of the most promising biomarkers to be used in the near future for HER2+ breast cancer, acording to the literature and the opinion of the healthcare professionals from HESE, and its correlation with this disease subtype.

Category	Biomarker	Correlation with HER2+ breast cancer	Reference
HER2 & HER3	HER2 levels	- Overexpression in all HER2+ breast cancer; HER2 expression is a continuum and therapeutic decisions should maybe be based on that (rather than positive/negative status).	[64], [78]
	HER2 mutations	- HER2 mutations are a resistance mechanism of anti-HER2 treatment. Evidence suggests that patients with these mutations have worse prognostics.	[66]
	HER2 heterogeneity	- HER2 expression and amplification may also show intratumoral heterogeneity. This heterogeneity may present in three main patterns: "clustered" type (with two distinct areas of the same tumor showing different HER2 status); "mosaic" type, displaying either diffuse intermingling of cells with different HER2 statuses; "scattered type",with positive and/or amplified cells dispersed within a negative tumor area. - The prevalence of HER2 genetic heterogeneity has been described in the range of 1–34%.	[64]
	HER3	- Important role in HER2+ breast cancer; Due to HER2/HER3 dimer, it is crucial in HER2-mediating signalling in breast cancer tumours; Development of anti-HER3 drugs is in process.	[64]
Gene expression	Intrinsic subtype (Prediction analysis of microarray 50 (PAM50))	- Several studies have established the role of intrinsic molecular subtype as a biomarker in the neoadjuvant setting for HER2+ breast cancer; There is also evidence that PAM50 subtypes may have implication on prognosis and treatment for metastatic HER2+ breast cancer patients.	[64], [79]
	PTEN	- Associated with the regulation of Pi3k-AKT pathway; low expression of PTEN have poor response to trastuzumab and shorter disease-free survival; In a recent study, the absence of PTEN IHC staining in tumor cells was associated with poor clinical outcome in HER2+ disease, although trastuzumab appeared to provide clinical benefit for patients lacking PTEN staining, suggesting that PTEN loss is associated with worse outcome rather than resistance to antibody treatment.	[80], [81]
DNA mutations	PIK3CA mutations and PI3K pathway inhibitors	- PIK3CA gene mutations are frequent in HER2+ breast cancer, occurring in 20% to 30% of patients; in the neoadjuvant setting, the presence of a PIK3CA mutation is associated with a lower rate of pathologic complete response (pCR) after chemotherapy and anti-HER2 treatment; Pi3K/Akt/mTOR pathway activation is one of the mechanisms of acquired resistance in HER2+ tumors. inhibitors are expected to overcome this resistance and to improve outcomes in previously refractory disease.	[65], [80], [78], [66], [64], [81], [82], [83]

Immune micro-environment	Tumour infiltrating lymphocytes (TILs)	- HER2+ breast cancer presents the high levels of TILs; Within HER2+ breast cancer,the level of TILs varies according to the molecular intrinsic subtype (high in basal-like and HER2 enriched tumours); In early HER2+ breast cancer, higher TILs have been associated with both increased likelihood of pCR after neoadjuvant therapy and with improved prognosis. - Within HER2+ breast cancer, the level of TILs varies according to the molecular intrinsic subtype, being higher in basal-like and HER2-enriched tumours. - TILs have also been shown a significant positive correlation with PD-L1 expression in both early and metastatic HER2+ breast cancer samples. -The International TILs Working Group has started standardizing evaluation of breast cancer TILs to be able to use this in clinical practice. - TILs might aid in the selection of candidates to immunotherapy.-The potential prognostic and predictive role of TILs in this setting needs further investigation in advanced disease.	[64], [66], [84], [85]
	PD-L1	- TILs show a significant positive correlation with PD-L1 expression in both early and metastatic HER2+ breast cancer samples. - HER2+ breast cancer is more likely to express PD-L1. - PD-L1 expression can be found either and non-simultaneously on tumour cells and on tumour infiltrating immune cells. In some cases, as in breast cancer, PD-L1 expression on tumour-infiltrating immune cells is a better predictor of response to immune checkpoint blockade than PD-L1 expression on tumour cells itself.	[64], [86], [87]
	FcγRs	- All currently approved monoclonal anti-HER2 antibodies are of the IgG isotype, comprising a crystalline fragment (Fc) linked to the antigen-binding fragments. The Fc domain interacts with Fc gamma receptors (FcγRs) expressed on a variety of immune cells; Some single nucleotide polymorphisms in the extracellular component of activating FcγRs have been associated with differential antibody binding affinity and antibody-dependent cell-mediated cytotoxicity (ADCC).	[64], [66]
	Liquid biopsy (ct-DNA, ct-miRNA, CTC)	- Liquid biopsy promises to evaluate the amplification of HER2 in a reproducible and non-invasive way which would allow the dynamic monitoring of efficacy and treatment decision-making.-Liquid biopsy is not limited to ctDNA and can mean evaluation of other potential markers (CTC and ct-miRNA had also been studied).	[66], [88]

5.2 Model structuring

5.2.1 Individual interviews with experts and decision-makers

With the results from the MEDI-VALUE project's Web-Delphi serving as a base, individual exploratory interviews with the two decision-makers and other six healthcare professionals from HESE were conducted, to assess which aspects are considered relevant to evaluate *in vitro* medical devices based on biomarkers for HER2+ breast cancer.

The results of the individual interviews are presented in two ways: the first concerning the individual responses of each health professional and the second relative to the distribution of responses, in absolute number, by the different aspects (Figure 5.2 and Figure 5.3, respectively).

Table 5.2: Distribution of individual responses, after the individual interviews, by healthcare professionals in each HESE service: Doctor 1 (D1) to Doctor 8 (D8), regarding the relevance of each aspect. [Response code: Cri (Critical), Fun (Fundamental), Com (Complementary), Irr (Irrelevant) and Dk (I do not know / I do not want to answer)]

Aspect	Hospital administration	Medical oncology	General surgery			Pathology		
	D1	D2	D3	D4	D5	D6	D7	D8
Specific characteristics of the medical device	Cri	Com	Irr	Irr	Cri	Cri	Cri	Com
Technical performance of the medical device	Fun	Com	Irr	Irr	Fun	Fun	Fun	Fun
Regulatory status of the medical device	Cri	Com	Com	Com	Cri	Fun	Fun	Fun
Sensitivity and specificity	Cri	Cri	Fun	Fun	Fun	Cri	Cri	Cri
Ease of use for healthcare professionals	Fun	Fun	Fun	Dk	Fun	Fun	Fun	Fun
Time between procedures and results	Com	Fun	Irr	Cri	Fun	Fun	Fun	Fun
Need for healthcare professionals' training	Fun	Fun	Irr	Fun	Com	Fun	Fun	Com
Healthcare professionals' learning curve	Com	Com	Fun	Dk	Com	Fun	Fun	Com
Healthcare professionals' exposure to physical or chemical agents	Cri	Com	Irr	Dk	Cri	Fun	Com	Fun
Healthcare professionals' workload	Fun	Com	Com	Fun	Com	Com	Fun	Fun
Patient's comfort	Com	Com	Irr	Irr	Fun	Fun	Irr	Cri
Connectivity	Com	Com	Irr	Fun	Fun	Fun	Irr	Com
Efficacy and/or clinical effectiveness	Fun	Cri	Cri	Fun	Fun	Com	Fun	Cri
Risk analysis	Fun	Com	Irr	Irr	Cri	Com	Irr	Cri
Adverse events for the patient	Cri	Cri	Irr	Irr	Cri	Irr	Fun	Fun
Quality of the available scientific evidence	Com	Cri	Fun	Cri	Cri	Fun	Cri	Fun
Target population	Com	Com	Fun	Com	Fun	Fun	Com	Cri
Disease impact	Com	Fun	Cri	Fun	Fun	Fun	Fun	Com
Patient-reported outcomes	Com	Fun	Irr	Com	Fun	Fun	Dk	Irr
Quality of life for the patient	Fun	Com	Fun	Fun	Fun	Fun	Fun	Com
Space for innovation in the organization	Cri	Com	Irr	Irr	Com	Irr	Fun	Fun
Clinical guidelines	Fun	Com	Irr	Com	Cri	Fun	Cri	Fun
Financing	Fun	Com	Com	Com	Com	Irr	Dk	Dk
Public health concern	Com	Fun	Fun	Fun	Fun	Fun	Com	Com
Impact on health system budget	Fun	Com	Com	Fun	Fun	Fun	Com	Irr
Equity	Fun	Fun	Com	Fun	Cri	Fun	Com	Com
Market competitiveness	Com	Com	Irr	Fun	Com	Com	Irr	Irr
Medical or technical complications for the patient	Fun	Fun	Irr	Irr	Cri	Irr	Fun	Cri
Agreement of key-actors in the adoption of the medical device	Cri	Cri	Fun	Com	Fun	Irr	Fun	Com
Environmental impact on the use of the medical device	Fun	Com	Irr	Dk	Fun	Irr	Com	Com
Efficiency	Fun	Fun	Fun	Fun	Fun	Fun	Fun	Com
Health system capacity	Fun	Fun	Fun	Fun	Fun	Fun	Fun	Fun
Medical device cost (including complementary equipment)	Fun	Fun	Irr	Cri	Fun	Fun	Com	Fun
Cost of the procedure without the medical device cost	Cri	Fun	Irr	Cri	Fun	Fun	Com	Com

Regarding the results from this initial phase, it is necessary to mention that, in some aspects, it was not possible to reach conclusions, since there was a polarity of results (*i.e.*, some healthcare professionals considered the aspect relevant, while others considered it irrelevant). These aspects were associated with a result "For further discussion" (Table 5.3), since they were re-analysed at an advanced stage, during the Workshop session with the decision-makers.

In addition to the direct responses given by the participants, some of them also added comments and provided feedback on some specific aspects, justifying the classifications they assign to the aspects. Besides, the participants also responded to the structured questions asked during the interview process. Both these answer types were of extreme importance since they provided a way to eliminate some aspects that were not relevant in this context.

Table 5.3: Distribution of responses (in absolute numbers) by the different aspects, and most common result. It should be noted that aspects where its relevance could be concluded from the responses (classified as either relevant or irrelevant, depending on the health professional interviewed) were analyzed and discussed later. In green, the results in which there is more than 50% agreement between the participants (5 or more participants); in yellow, the most common result for the aspect, in cases where there is less than 50% agreement. [Response code: Cri (Critical), Fun (Fundamental), Com (Complementary)]]

Aspect	Answers (in absolute number)					Result
	Critical	Fundamental	Complementary	Irrelevant	I do not know/ I do not want to answer	
Specific characteristics of the medical device	4	0	2	2	0	For further discussion
Technical performance of the medical device	0	5	1	2	0	For further discussion
Regulatory status of the medical device	2	3	3	0	0	Fun/Com
Sensitivity and specificity	5	3	0	0	0	Cri
Ease of use for healthcare professionals	0	7	0	0	1	Fun
Time between procedures and results	1	5	1	1	0	Fun
Need for healthcare professionals' training	0	5	2	1	0	Fun
Healthcare professionals' learning curve	0	3	4	0	1	Com
Healthcare professionals' exposure to physical or chemical agents	2	2	2	1	1	Cri/ Fun/ Com
Healthcare professionals' workload	0	4	4	0	0	Fun/ Com
Patient's comfort	1	2	2	3	0	For further discussion
Connectivity	0	3	3	2	0	For further discussion
Efficacy and/or clinical effectiveness	3	4	1	0	0	Fun
Risk analysis	2	1	2	3	0	For further discussion
Adverse events for the patient	3	2	0	3	0	For further discussion
Quality of the available scientific evidence	4	3	1	0	0	Cri
Target population	1	3	4	0	0	Com
Disease impact	1	5	2	0	0	Fun
Patient-reported outcomes	0	3	2	2	1	For further discussion
Quality of life for the patient	0	6	2	0	0	Fun
Space for innovation in the organization	1	2	2	3	0	For further discussion
Clinical guidelines	2	3	2	1	0	Fun
Financing	0	1	4	1	2	Com
Public health concern	0	5	3	0	0	Fun
Impact on health system budget	0	4	3	1	0	Fun/ Com
Equity	1	4	3	0	0	Fun
Market competitiveness	0	1	4	3	0	For further discussion
Medical or technical complications for the patient	2	3	0	3	0	For further discussion
Agreement of key-actors in the adoption of the medical device	2	3	2	1	0	Fun
Environmental impact on the use of the medical device	0	2	3	2	1	For further discussion
Efficiency	0	7	1	0	0	Fun
Health system capacity	0	8	0	0	0	Fun
Medical device cost (including complementary equipment)	1	5	1	1	0	Fun
Cost of the procedure without the medical device cost	2	3	2	1	0	Fun

General comments to the aspect

In addition to specific comments made on certain aspects, some general comments made by the healthcare professionals regarding the questions asked during the interviews (namely the answers provided to questions 2., 3. and 4. from Chapter 4, Section 4.2.2) were also made.

Regarding question 2.: *"Do you consider that the aspects presented are easy to understand, concerning the knowledge you have in the area? If not, what do you think could be improved?"*, most professionals referred that the set of aspects was too general and that a focus should be made to specify it for the context of biomarkers for HER2+ breast cancer. In addition, a healthcare professional reported that most aspects were very technical and that a greater focus on the clinical part would also be important. Regarding this question, there was some confusion among the healthcare professionals when assessing aspects, since some professionals considered the initial list of aspects to be very general for the specific case of *in vitro* tests based on biomarkers for HER2+ breast cancer. This fact can be considered as a limitation in the individual exploratory interviews and the consequent classification of the aspects concerning their relevance.

Regarding question 3. *"Within the mentioned aspects, do you consider that there is a missing aspect that should be included?"*, all professionals considered the list to be quite exhaustive and complete, touching on all the points that they considered relevant. Note that the initial list of aspects was considered too long, according to some healthcare professionals.

Finally, for question 4. *"Do you approve the list of potential biomarkers for HER2+ breast cancer?"*,

all of the eight participants provided their feedback. In fact, one of them showed particular enthusiasm for: HER2 mutations, PIK3CA mutations and PI3K pathway inhibitors, PD-L1, and Liquid biopsy. Two of the health professionals did not feel comfortable giving their approval regarding the list sent, as they considered that they did not have enough information about these emerging biomarkers (they work with those already established by international guidelines and that are already used in clinical practice, at HESE, and consider that only about these can have a reasoned opinion). Despite these two professionals, all the others were comfortable in approving the list.

5.2.2 Workshop

Workshop Part I

During the first part of the Workshop, it was clarified with the decision-makers which was the focus of the study. After some discussion, the participants decided that, despite being an ambitious focus, it was possible and appropriate to compare the different biomarkers with each other, despite the distinct functions that they might present (of diagnosis, prognosis, and prediction of drug/therapy response). Thus, the initial and provisional list of biomarkers (previously approved by the group of eight health professionals during the stage of the individual interviews) was once again analysed and reformulated, resulting in the final list, presented in Table 4.1.

After the focus was defined, it was then possible for the decision-makers to classify the aspects according to their relevance, for the context of HESE. After the classification, and out of the 34 initial aspects, the list was reduced and only 18 aspects were considered relevant at this stage (Table 5.4).

Table 5.4: The list of the 18 aspects considered relevant by the two healthcare professionals and its respective classification, after Workshop Part I. Note that the aspects considered irrelevant were eliminated at this stage.

Aspect	Classification
Sensitivity and specificity	Fundamental
Ease of use for health professionals	Fundamental
Time between procedures and results	Fundamental
Need for health professionals training	Fundamental
Health professional's learning curve	Complementary
Exposure of the health professional to physical or chemical agents	Fundamental
Workload for health professionals	Fundamental
Efficacy and/or clinical effectiveness	Critical
Quality of the available scientific evidence	Complementary
Disease impact	Fundamental
Space for innovation in the organization	Fundamental
Clinical guidelines	Complementary
Public health concern	Complementary
Impact on health system budget	Fundamental
Agreement of key actors in the adoption of the medical device	Fundamental
Efficiency	Fundamental
Health system capacity	Fundamental
Medical device cost (including complementary equipment)	Fundamental

During the aspect classification process, some comments were made that allowed the decision-makers to justify some of the aspects irrelevance in this context, while other comments were pertinent to build the performance scales, at a later stage of the project. The main comments made were regarding the following aspects:

- Time between procedures and results: aspect considered relevant as it depends on the allocation of personnel (if 1 technician is needed, 2 technicians are needed, etc.);

- Healthcare professionals' exposure to physical or chemical agents: considered a relevant aspect, since the professional is in contact with reagents, with different hazards depending on the biomarker;

- Patient's comfort: they do not consider it relevant that the patient's satisfaction is altered by using a certain biomarker, to the detriment of another;

- Quality of the available scientific evidence: considered a somewhat relevant (complementary) aspect since there is still little information available associated with the biomarkers evaluated;

- Disease impact: aspect considered relevant, since different groups may respond differently (and therefore, there are different outcomes depending on the HER2+ subtype);

- Financing: considered an irrelevant aspect in the context, since the hospital pays the same amount for any type of patient with breast cancer, regardless of the tests performed and the therapy adopted. The hospital then receives a certain amount, so the return is the same. However, it is important to optimize costs;

- Equity: aspect considered important in a study context. It does not apply in the context of these specific biomarkers;

- Cost of the procedure without the medical device cost: considered as an irrelevant aspect, since the high price of these tests is associated with the reagents (and therefore, directly related to the medical device itself and not with auxiliary elements).

At the time of the Workshop Part I and as mentioned in Chapter 4, it was not possible to address some concerns (overlaps and redundancies between the selected aspects; development of descriptors of performance and scales for the aspects). Thus, these concerns were addressed in the following session: Workshop Part II.

Workshop Part II

The first activity proposed to the two decision-makers for Workshop Part II was to identify potential overlaps, redundancies, and inter-dependencies between the different aspects. For that, a value tree was presented, in which the aspects were organized in different dimensions of evaluation. It was also mentioned during this exercise that the costs associated with the biomarker were only going to be considered afterwards, following a logic of *Value for Money*.

At this stage, the decision-makers were asked to analyze the value tree and, if necessary, to suggest changes in its organization. They referred that the evaluation dimensions considered and the aspect organization were coherent, which allowed proceeding to the validation/discussion in greater detail regarding some characteristics of the value tree.

Then, the performance scales developed, based on the available literature and scientific evidence, were presented to the decision-makers, for each of the aspects that were represented in the value tree. With the following exercise, it was intended that the decision-makers gave their opinion, based on their experience in the area, regarding the performance scales previously built. As mentioned in Chapter 4 section 4.2.2, besides the input provided during Workshop Part II, it was also necessary to speak with other two healthcare professionals (directors of the Pathology and Medical oncology services, at HESE) to have further insight to build the value tree and the performance scales. As a result, some of these scales did not undergo any changes and remained the same as those initially proposed. However, in the majority, proposals for improvement and alteration were mentioned by these two

healthcare professionals, including grouping aspects due to redundancies between them; removing some aspects; and consider some aspects only at a later stage, from the perspective of *Value for Money*, and for a future cost-benefit analysis.

It is important to stress that the contribution of these two professionals from the Pathology and Medical oncology services from HESE, expressed after Workshop Part II, was of extreme importance, since it provided further insight and their inputs were an excellent addition to the results obtained from the Workshop (both Part I and Part II).

After this, it was possible to decide which were the aspects (Table 5.5) that were going to be considered for the evaluation of the biomarkers' options (and thus, be considered as criteria), and to build the descriptors and performance scales for the criteria (Table 5.6), as well as a final value tree (Figure 5.1).

Table 5.5: Criteria, criteria relevance and description.

Criteria	Relevance	Description
Clinical relevance	Critical	Biomarker's capacity to identify subtypes of the disease, and its expected impact on clinical practice.
Existence of clinical guidelines and quality of evidence	Complementary	Robustness of the evidence sources and quality of the evidence regarding the target population and the intended clinical course, and recommendation of the biomarker by national and/or international clinical guidelines.
Usability for the healthcare professional	Complementary	Extent to which the procedures for using the biomarker are defined and clear, and the result is easy to interpret. The need for training the healthcare professional, and the learning curve are also considered.
Impact on the form of work and workload	Fundamental	Extent to which the biomarker can be implemented without drastic changes in the current form of work and in the workload of healthcare professionals.
Agreement of key-actors in the decision	Fundamental	Agreement between key-actors (stakeholders) in the hospital context, that is, the extent to which the adoption of the biomarker is aligned with common objectives.
Public health concern	Complementary	Ability to use the biomarker as a tool for diagnosis, prognosis, or prediction of drug/therapy response.

Table 5.6: Descriptors of performance for the criteria (with different levels, L), and its *neutral* and *good* reference levels.

Criteria	Descriptors of performance
Clinical relevance	**L3-** It is expected the use of the biomarker in clinical practice to be very relevant, since it has the potential to identify different subtypes of the disease and, consequently, allow more effective treatments. It is already available for use in certain cases, in clinical settings. **L2-** It is expected the use of the biomarker in clinical practice to be very relevant, since it has the potential to identify different subtypes of the disease and, consequently, allow more effective treatments. It is not yet available for use in a clinical setting. **[GOOD]** **L1-** It is expected the use of the biomarker in clinical practice to be relevant, with an indirect impact on the treatment of the disease since it is not a crucial biomarker to identify different subtypes of the disease. There are other biomarkers already used in clinical practice that allow for proper diagnosis and treatment. However, it can be used as a complementary instrument. It is not yet available for use in a clinical setting. **[NEUTRAL]**
Existence of clinical guidelines and quality of evidence	**L3-** Biomarker approved at least by one international guideline (*e.g.*, ASCO, NCCN, ESMO, . . .) for use in HER2+ breast cancer in a clinical setting. **[GOOD]** **L2-** Biomarker approved at least by one local or national scientific group (without recognition by international or national guidelines). **L1-** Biomarker not yet recognized by any international guideline, but with convincing clinical evidence demonstrated in clinical trials. **[NEUTRAL]**
Usability for the healthcare professional	**L4-** The test based on the biomarker requires specific background (through in-house training - training through a hospital colleague) to be handled. The result is easy and simple to interpret (peak proficiency quickly reached by the professional since first use). **[GOOD]** **L3-** The test based on the biomarker requires specific background (through external training - credited training) to be handled. The result of it is easy and simple to interpret (peak proficiency quickly reached by the professional since first use). **[NEUTRAL]** **L2-** The test based on the biomarker requires specific background (through in-house training - training through a hospital colleague) to be handled. However, the result is sometimes difficult to interpret (peak proficiency slowly reached by the professional since the first use). **L1-** The test based on the biomarker requires specific background (through external training - credited training) to be handled. However, the result is sometimes difficult to interpret (proficiency peak slowly reached by the professional since the first use).
Impact on the form of work and workload	**L3-** The implementation of this biomarker in clinical practice will have no impact on the form of work or on the health professional's workload. **[GOOD]** **L2-** The implementation of this biomarker in clinical practice will have no impact on the health professional's workload. However, there will be changes on the form of work. **L1-** The implementation of this biomarker in clinical practice will have an impact on the form of work and the health professional's workload. **[NEUTRAL]**
Agreement of key-actors in the decision	**L5-** Unanimity: all key-actors agree with the adoption of the biomarker (100%). **L4-** Quasi-unanimity: almost all (around 90%) the key-actors agree with the adoption of the biomarker. **[GOOD]** **L3-** Qualified majority: a significant number of key-actors (around 75%) agree with the adoption of the biomarker. **[NEUTRAL]** **L2-** Simple majority: more than half of the key-actors (around 51%) agree with the adoption of the biomarker. **L1-** Non-agreement, among the key-actors in the adoption of the biomarker (below 51%).

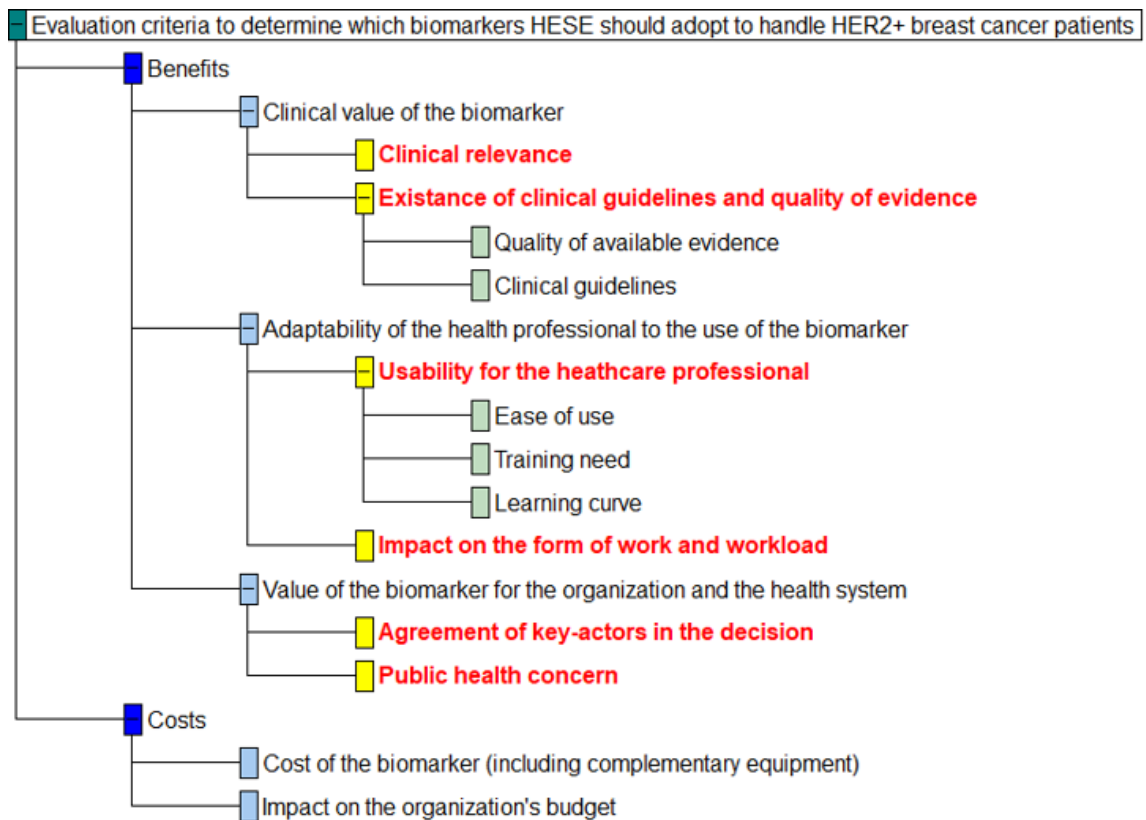|	
Public health concern	**L3-** The use of the biomarker is of high interest to public health, since it can be used to: • Diagnosis, prognosis, and prediction of drug/therapy response for the case of HER2+ breast cancer. **L2-** The use of the biomarker is of interest to public health, since it can be used to: • Diagnosis and prognosis or • Prognosis and prediction of drug/therapy response for the case of HER2+ breast cancer. **[GOOD]** **L1**- The use of the biomarker is of interest to public health, since it can be used to: • Diagnosis or • Prognosis or • Prediction of drug/therapy response for the case of HER2+ breast cancer. **[NEUTRAL]**

Figure 5.1: Final value tree. The dark blue nodes represent the two areas of concern, while the light blue nodes correspond to the different categories in which the criteria (in red and in yellow nodes) are organized. The grey nodes represent elementary points of view that constitute the criteria [source: M-MACBETH software]

5.3 Model building

5.3.1 Decision Conference Part I

As previously referred to in Chapter 4, two activities were conducted during this phase: the construction of value scales for each criterion and the weighting of the criteria.

Regarding the first activity, it was asked for the decision-makers to make pairwise comparisons, using the MACBETH qualitative judgements, between the levels of each criterion. For that, comparisons between two levels of the descriptor of performance were presented and it was asked the decision-makers to judge the difference in attractiveness between the pairs. By doing so, it was possible to construct a judgement matrix for each criterion (consult Appendix B) that would result in the final value scales, after some adjustment were made, for a final approval by the decision-makers (Figure 5.2).
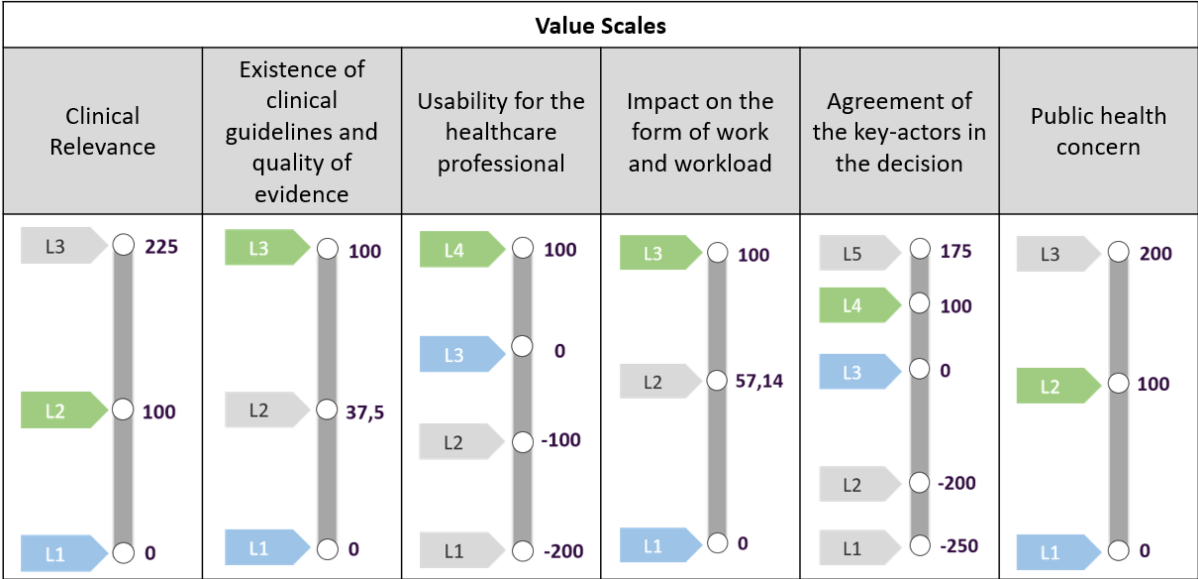
Figure 5.2: Value scales obtained for each of the criteria, using the MACBETH technique, and after adjustments and approval by the decision-makers. Note that, for each criterion, the level (L) represented in green corresponds to the *good* reference level (score 100), while the level (L) represented in blue corresponds to the *neutral* reference level (score 0). [source: WisedOn software - adapted figure, not represented to scale]

Afterwards, the second and last activity of this session was to assign the weights to the different criteria. To do so, the swings between the *neutral* and *good* levels were ranked, by the decision-makers (Figure 5.3). With this ranking, the criteria were organized according to its importance, meaning that the swing, from *neutral* to *good*, associated with the criterion placed first (or on top) was more important for the decision-makers than the other swings, associated with the other criteria, placed last (or at the bottom).

After this organization, it was then possible to compare the swings (with the *neutral* reference level or with another criterion), to build the weighting matrix of judgements, that provides the relative weights for each criterion (consult Appendix C). Following the 21 questions asked to fill every cell of the matrix, the weights were obtained.

The obtained weights were then presented to the decision-makers and the results were discussed. It was referred that the weights should be rounded up to the nearest unit, for a simpler and evidence presentation of the results. Thus, this suggestion was considered, and the final result is presented in Figure 5.4.

Not yet recognize by any guideline	→	Approved at least by 1 international guideline	Existance of clinical guidelines and quality of evidence (complementary)
Indirect impact, is not yet available in clinical settings	→	Very relevant, but it is not yet available in clinical settings	Clinical relevance (critical)
Diagnostic OR Prognostic OR Prediction of drug/therapy response	→	Diagnostic and prognostic OR Prognostic and prediction of drug/therapy response	Public health concern (complementary)
Qualified majority (around 75%)	→	Quasi-unanimity (around 90%)	Agreement of key-actors in the decision (fundamental)
Impact on the form of work and workload	→	No impact on the form of work and workload	Impact on the form of work and workload (fundamental)
External training, result is easy and simple to interpret	→	In-house training, result is easy and simple to interpret	Usability for the healthcare professional (complementary)

Figure 5.3: Different criteria ranked from the most important one (at the top) to the least important one (at the bottom). The comparison was made between the *neutral* reference level (abbreviated in the left, in grey) and the *good* reference level (abbreviated in the middle, in green). [source: WisedOn software]
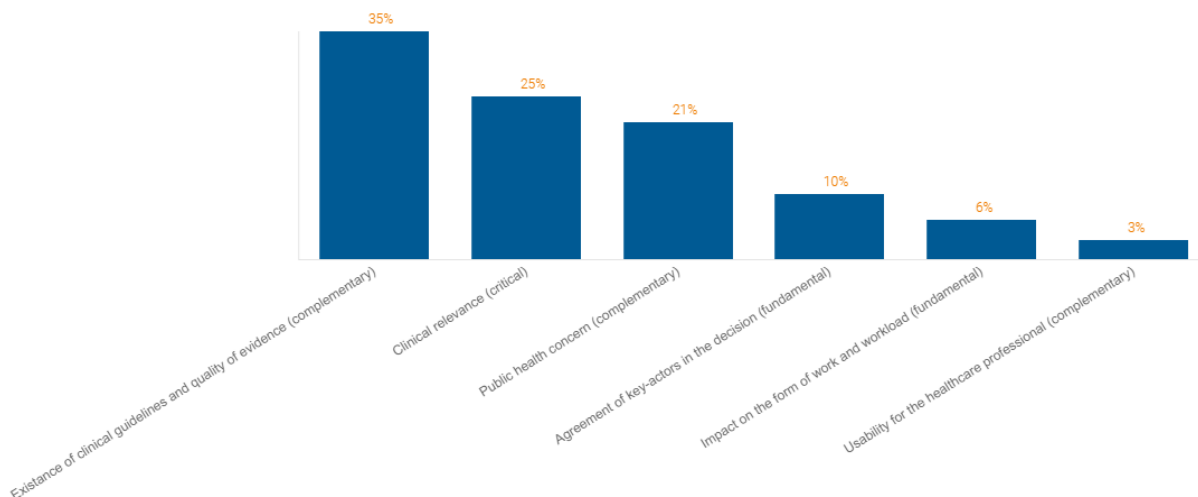
Figure 5.4: Weights associated with each criteria. [source: WisedOn software]

Besides, it is also presented in Table 5.7 the partial scores associated with each biomarker, as well as the global scores of each biomarker option (consult Appendix D for the visualization of the biomarkers' performance in the different criteria, according to its descriptor of performance).

Table 5.7: Partial score for the different biomarkers in each criterion and global score for each of them, ranked from higher global score to lower global score. Note that there is a tie between two biomarkers (T8), as they have the same global score. [source: WisedOn platform - adapted].

Option (biomarker)	Partial score in each criterion						Global score (#ranking)
	Clinical relevance (25%)	Existence of clinical guidelines and quality of evidence (35%)	Usability for the healthcare professional (3%)	Impact on the form of work and workload (6%)	Agreement of key-actors in the decision (10%)	Public health concern (21%)	
PD-L1	225	100	100	57.14	175	0	115.18 (#1)
Liquid biopsy (ctDNA, ct-miRNA, CTC)	100	37.5	0	0	100	200	90.13 (#2)
PIK3CA mutations and PI3K pathway inhibitors	225	0	100	57.14	175	0	80.18 (#3)
TILs	100	37.5	100	57.14	100	100	75.55 (#4)
Intrinsic subtype (PAM50)	225	0	0	0	-250	100	52.25 (#5)
HER2 levels	0	0	100	57.14	-250	200	23.43 (#6)
HER2 mutations	100	0	0	0	-250	100	21 (#7)
HER2 heterogeneity	0	0	0	0	-250	100	-4 (T8)
HER3	0	0	0	0	-250	100	-4 (T8)
PTEN	0	0	0	0	-250	0	-25 (#10)
FcγRs	0	0	-200	0	-250	0	-31 (#11)

5.4 Model testing and validation

5.4.1 Decision Conference Part II

In the last session with the decision-makers, the overall scores for the biomarkers were presented (Figure 5.5).
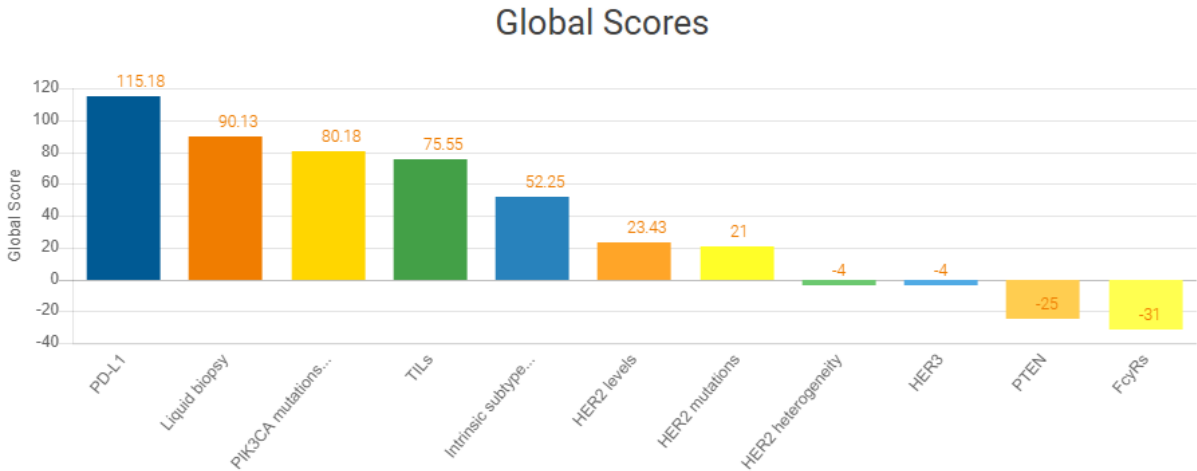
Figure 5.5: Global scores for each of the 11 biomarkers. [source: WisedOn software]

Regarding the results presented in Figure 5.5, the decision-makers considered the biomarkers scores expected. They also considered the 3 biomarkers with the highest scores as the most significant and most discussed according to their experience (PD-L1, Liquid Biopsy, PIK3CA mutations and PI3K pathway inhibitors). In addition to these 3 biomarkers, the 4th biomarker with the highest score (TILs)

was also considered by the participants as deserving of attention. In addition to this analysis, a profile comparison was also carried out between 2 pairs of biomarkers: PD-L1 / HER2 mutations, and PIK3CA mutations and inhibitors of the PI3K pathway / PTEN, using the WisedOn software, to analyse, in further detail, in what aspects did the scores of these pairs of biomarkers differentiated.

Sensitivity analyses were also conducted to understand, to which extent, the criteria weights would interfere with the biomarkers' scores and rank (example in Figure 5.6). Multiple sensitivity analyses were conducted, and in some cases, more than one for each criterion, since in some examples only a subgroup of biomarkers was analysed.
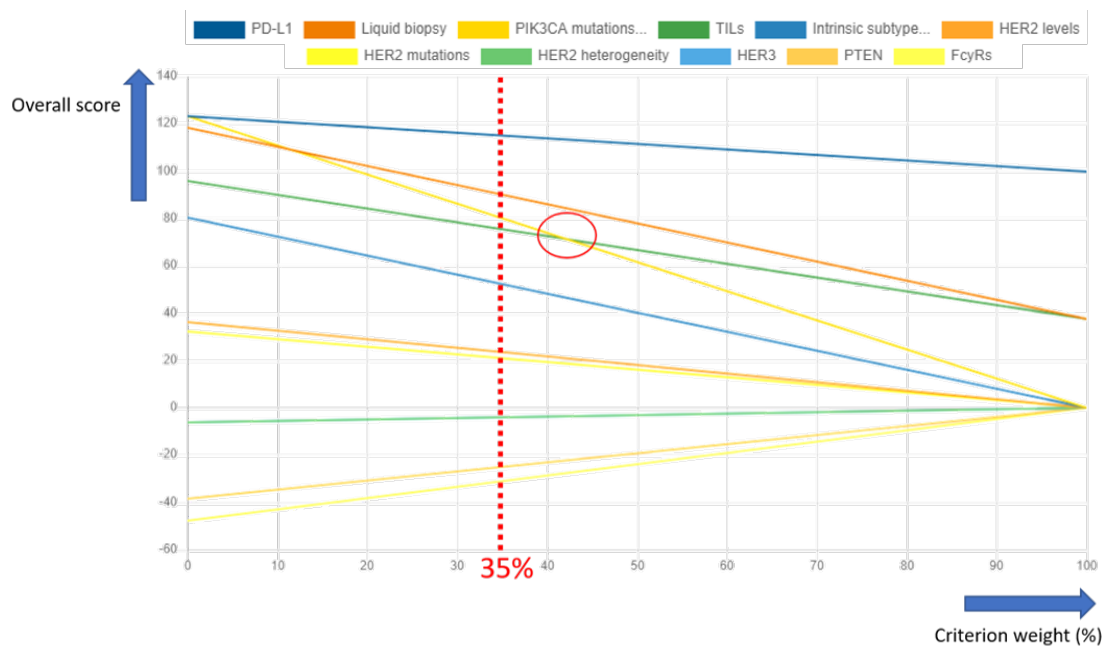
Figure 5.6: Example of a sensitivity analysis performed for the criterion *Existence of clinical guidelines and quality of evidence*, with a relative weight 35% (represented by the red dashed vertical line). Note the intersection delimited by the red circle represents a change in the ranking of biomarkers (TILs and PIK3CA mutations and inhibitors of the PI3K pathway) close to the current weight of the criterion - in case the relative weight of the criterion increased to values around 41-44%. [source: WisedOn software - adapted figure]

Finally, decision-makers were asked to provide feedback on the study in which they had participated. As already mentioned, questions were asked for the last three phases of the methodology: model structuring, model building, and model testing and validation, the phases in which the decision-makers participated.

Regarding the model structuring phase, the following questions were asked to the decision-makers:

- Was the objective of the case study clear from the beginning?
- Which were the biggest difficulties that arose during this phase?

The decision-makers provided positive feedback concerning this phase. They commented that the case study objective was clear since the beginning, despite the change of focus at the start of the project: initially, the study was more generalized, for the emerging biomarkers for all breast cancer subtypes; only after some discussion with the healthcare professionals at HESE the focus was tightened for the HER2+ breast cancer subtype. Nonetheless, they also affirmed that this did not work as an impediment to understanding the objective of the study, in broader terms.

They also added that, during the model structuring phase, the biggest difficulty that they had to face was the construction of performance scales, due to the reduced information available for emerging

biomarkers and the high number of biomarkers' options that were considered. They also declared that maintaining objectivity during the entire duration of the study was a struggle. This concern is, of course, expectable, since the model deals with multiple individual's concerns, opinions, and points of view, being very demanding for the decision-makers to always work objectively.

Concerning the model building phase, the decision-makers were confronted with the following questions (similar to the ones asked for the previous phase):

- Was it easy/intuitive to answer the questions asked during the Decision Conference Part I?

- What were the biggest difficulties that arose during the Decision Conference Part I?

Concerning the first query, the decision-makers considered that it was easy to respond to the questions posed during the Decision Conference Part I, held in the WisedOn software. They referred that the software was extremely intuitive and recognised it to be a powerful tool, very useful during the development of this study.

Once again, as for the difficulties that emerged during this phase, the decision-makers stated that it was not always easy to remain objective, due to preconceived notions and opinions they had.

Lastly, the questions concerning the last phase:

- Was it easy/intuitive to analyse the results obtained and carry out the sensitivity analyzes associated with each aspect?

- Are the results obtained relevant and do they have the potential to inform the adoption of biomarkers in HESE?

- If it were possible to carry out a cost-benefit analysis, would it be relevant considering the objective of the case study?

- What improvements do you consider relevant to the proposed methodology, in a future project, with a view to its effective application in a real context?

- Do you have any final thoughts or comments that you want to share?

The decision-makers also gave positive feedback: they considered that the results obtained were presented in a clear and simple way, the same with the sensitivity analysis. The results were also considered relevant and could serve as a basis for the future adoption of biomarkers in HESE. However, for this, the costs associated with the biomarkers must be taken into account as well, which leads us to the next question asked.

Regarding the cost-benefit analysis: the decision-makers stated that it is always important to know the associated costs, especially in a healthcare organization (such as HESE), since the cost is crucial during the adoption decision phase. However, at the initial level where the study is located, "cost-benefit analysis is important, but it is not the most important", as one decision-maker stated. They stressed that once the biomarkers with the most interest have been identified and the benefit has been ascertained, it would then be interesting to carry out this analysis, in a future phase and/or study.

The decision-makers also considered that, in a next study, it will be important to limit the number of biomarkers under analysis, so that the study can be more focused. In their opinion, this first study was important, to identify the biomarkers with the highest value and to be able to filter the information for a second study, which, using more evidence, has a stricter focus, only for the most promising biomarkers. Furthermore, in the current study, the decision-makers also considered the goal of opting for a study concerning emerging biomarkers was ambitious since there is still little information available in the literature when compared to the existing information for biomarkers already used in clinical practice. However, this scope was also considered to be important, to centre the focus only on the most relevant biomarkers.

Finally, and according to the results obtained, there was a biomarker that caught the attention of the decision-makers: PTEN appears as a biomarker with a very low (and negative) score. The decision-makers showed some surprise concerning its score. However, after analysing their profile (during Decision Conference Part II) and taking into account the existing information about this biomarker and its valuation in different aspects, they understand that a focus on this biomarker is not yet a priority.

In the following chapter, Chapter 6, a reflection regarding the methodology used, as well as the main challenges and limitations encountered will be done.

Chapter 6

Discussion

In this chapter, a discussion concerning the methodology developed will be presented. In addition, a reflection regarding the strengths, challenges, and limitations associated not only with the methodology, but also with this thesis as a whole, will also be done.

6.1 Used methodology

Despite being applied in multiple health-related contexts, MCDA, and more specifically the MACBETH approach is still not used in many studies associated with the adoption of biomarkers, for clinical use and application. This project, as part of the MEDI-VALUE project, tries to fill the gap, by providing a structured methodology to evaluate emerging biomarkers for future use in the clinical context.

This methodology can be applied, not only for these specific biomarkers but can also be used in other contexts, to analyse different emerging biomarkers for other diseases.

For the development of this methodology, a socio-technical approach was used, which involved four main phases: evidence synthesis, where a rapid review was conducted, as well as a synthesis of stakeholders and experts' option results from a previous Web-Delphi; the model structuring phase, using individual interviews and a Workshop; the model building and, lastly, the model testing and validation, both through a Decision Conference held with the decision-makers. This allowed to reach a final multicriteria evaluation model, within the scope of HTA, more specifically, HB-HTA, to assist decision-making in HESE regarding emerging and promising biomarkers that should be adopted to be used to optimize the current diagnosis, prognosis, and prediction of drug/therapy response in HER2+ breast cancer.

The resulting multicriteria model presents six criteria that were considered relevant to assess and evaluate the benefit of emerging biomarkers: *Clinical Relevance*, *Existence of clinical guidelines and quality of evidence*, *Usability for the healthcare professional*, *Impact on the form of work and workload*, *Agreement of key-actors in the decision* and *Public health concern*. After the evaluation of the emerging biomarkers, the four which presented the higher scores - PD-L1, Liquid biopsy, PIK3CA mutations and PI3K pathway inhibitors, and TILs were considered by the decision-makers as the ones who hold the greatest promise for the diagnostic, prognostic, and prediction of drug/therapy response in HER2+ breast cancer.

6.2 Strengths and difficulties associated with the methodology

Despite being created and used for the specific context of the HER2+ breast cancer subtype, one significant advantage of the model is that it can be reused, whenever needed, to assess other

emerging and new biomarkers, that may be of interest for HESE, not only for breast cancer but also for other diseases.

This structured methodology allowed the participants to always understand in which phase of the process they were in, as well as having a general view of the project, important on long projects such as this one, reasonably spaced in time (contact with HESE from May 2020 to April 2021). Moreover, despite some initial difficulties felt by the participants, they gradually started to think in a *value-focused* way, also reflected on their participation and commitment to the project.

Regarding the use of MCDA, specifically MACBETH, it was very advantageous for this project: the development of two distinct phases to choose the criteria, with multiple participants (in the individual interviews, with eight healthcare professionals, and the Workshop, with two decision-makers) allowed to reach a final set of criteria, according to the participants, to evaluate emerging biomarker for the HER2+ breast cancer subtype. The search for possible inter-dependencies and redundancies among the aspects enabled a final list of criteria that were independent of one another, an important characteristic in MCDA methodology. Another important characteristic of the methodology, more related to MACBETH, corresponds to the request of qualitative judgements, instead of quantitative information, from the decision-makers. This showed to be more intuitive for the participants since it is difficult in this type of model to provide quantitative data. The fact that it was possible to allow a classification window while providing the qualitative judgements in cases of hesitation, was also pointed as an asset in the process. The sensibility analyses were also positively commented, as an important way to accept and validate the results obtained.

The use of the WisedOn software was of extreme importance during the contact with the decision-makers, in the Decision Conference sessions. It is a very simple tool, based on the M-MACBETH software and, thus, it involves the same protocol. Due to its modern and user-friendly design, it allowed to easily collect the qualitative judgements, as well as the visualization and exploration of the results.

The feedback asked to the decision-makers in the last phase of the methodology was also of extreme relevance, as it proved to be a powerful tool to understand how the process unfolded, to identify which were the positive aspects associated, what went wrong, and even what could be improved in a future work of this nature.

Lastly, it is expected that the approach developed in this thesis can contribute to MCDA in HTA, more specifically, related with an early stage of HTA, *i.e.*, managing emerging medical devices at a hospital level. It is expected that this work can provide the required foundations to serve as a base for other works and studies in this area.

Regardless, it is still important to pinpoint some of the difficulties that were felt, and some limitations associated with the development of this project.

The COVID-19 pandemic that took the world by surprise, days after the initiation of the project was one of the most serious problems since it was present from its beginning to its conclusion. It was difficult to establish contact with HESE, in such unusual times, due to the enormous pressure to which they were subjected. Thus, it was necessary to conform with remote contact, by organizing phone calls and video conferences with the participants when they were available. As a direct consequence, the process stretched in time and, in some periods, it was exhausting to the participants.

Another difficulty was associated with the biomarker options. Initially, the focus was broader, for breast cancer in general. Then, after some discussion with the decision-makers, it was decided that only biomarkers for the HER2+ breast cancer subtype were going to be evaluated. Besides, the fact that the options were emerging biomarkers also limited the study: it was hard to find information since most of them are still not used in clinical practice, and others, despite already being used, are not yet applied in the context of HER2+ breast cancer. Lastly, and as a direct consequence of working with emerging biomarkers relies on the fact that it was not possible to conduct a cost-benefit analysis, as

initially desired, due to lack of information regarding costs and commercial prices of the options.

Despite being initially planned, it was not possible to conduct a cost-benefit analysis for the biomarkers' options. This was due to the fact that the majority of the options under study are still not used in clinical practice. However, the ones that are already implemented, are used for other breast cancer subtypes or other types of cancers and are not used for the HER2+ breast cancer subtype. Another important note is associated with the classification of the aspects, according to its relevance (Critical, Fundamental, or Complementary) that was considered throughout the methodology. It was asked for the aspect relevance, not only to the initial group of experts during the individual exploratory interviews but also to the decision-makers during the Workshop (Part I and Part II), to assess if the aspects associated with higher relevance during the model structuring phase would reflect aspects with higher weights, in the building phase. Despite the initial impression that they would be correlated, the results showed that aspects initially considered by the decision-makers as Complementary had higher weights than ones considered Fundamental. In fact, one of the aspects classified as Complementary (*Existence of clinical guidelines and quality of evidence*) had a higher weight associated than the one classified as Critical (*Clinical relevance*) - consult Figure 5.4 - which goes against the initial conjecture that was expected.

Chapter 7

Conclusion

This thesis, inserted in the MEDI-VALUE project, had the main objective of developing an innovative methodology to assess and evaluate a medical device, in the scope of HTA. HESE, due to its renown in dealing with oncological diseases, showed great interest in assessing emerging biomarkers for breast cancer disease, referenced by the literature, but not yet recommended by international guidelines nor used in clinical practice. Due to the multiple existing disease subtypes, a focus was placed on the HER2+ breast cancer subtype, which is very aggressive and is associated with an unfavorable patient prognostic.

Thus, a socio-technical approach that enabled the multicriteria evaluation of emerging biomarkers for HER2+ breast cancer was developed and implemented. This approach was innovative, as different tools were integrated throughout the process. The literature review on HTA, more specifically HB-HTA, allied with the current needs for biomarkers in the breast cancer field, permitted an extended understanding of the current challenges that are being faced in this area. In addition, the literature review in the area of MCDA allowed for this multi-methodology, in which a model for evaluation of emerging biomarkers was constructed. Besides, the participation of multiple healthcare professionals allowed the development of an inclusive, representative and valid model. In the first step of the methodology, a rapid review procedure was conducted, which provided crucial information regarding which biomarkers to assess and evaluate. In parallel, Web-Delphi results from a MEDI-VALUE project were used and also integrated into the methodology, regarding the identification of relevant aspects for the evaluation of medical devices, for the particular case of *in vitro* tests based on biomarkers and implanted medical devices. During the model structuring phase, individual interviews, as well as a Workshop were conducted with healthcare professionals from HESE, which allowed to simplify and identify a list of criteria and to construct the descriptors of performance and reference levels for each criterion. The participation of healthcare professionals from multiple backgrounds during the individual interviews was extremely advantageous to this project since it allowed this project to be more reliable, contributing to its validation and credibility. Lastly, the third and fourth phases of the methodology were conducted through two Decision Conferences, allowing for the construction of the value scales, and definition of weights (during the model building phase); and examination, and consequent model approval through sensitivity analysis (during the model testing and validation phase).

After the mentioned phases, it was possible to obtain an innovative, original and consolidated model that enabled the multicriteria evaluation of emerging biomarkers, a tool that can be used in HESE, providing further knowledge and insights regarding emerging biomarkers for the HER2+ breast cancer subtype. Besides, and despite being implemented for biomarkers associated with a specific breast cancer subtype, that have the potential to integrate *in vitro* tests, the model can be reused and adapted for any type of biomarker, for any type of disease (not necessarily in the area of breast cancer) and for context that can differ from the one of HESE. This is undoubtedly advantageous since this MCDA model

can be reused and expanded to other areas of HTA (more specifically for HB-HTA) to assess any other type of health technology.

It is expected that this thesis can provide further knowledge and insights regarding the growing field of HTA, specially in hospital context (HB-HTA), and MCDA so that it can serve as a support for other future studies in the area.

7.1   Future work

As already mentioned, this thesis was inserted in the MEDI-VALUE project and served as a pilot study, since it will become a support for the development of other MEDI-VALUE works. The future developed models will not only be in the field of biomarkers but will also assess other medical devices, namely cardioverters and esophageal/gastric prosthesis, that will be executed with other MEDI-VALUE partners.

In the future, and as a continuity of the work developed, it would be interesting to also perform a *Value for Money* analysis, which was impossible to perform at this stage, due to lack of information regarding the emerging biomarker's costs. Only with such analysis would it be possible for HESE to adopt a health technology since the economical component is of extreme importance in a hospital adoption process. Besides, using this type of analysis, it would also be possible to eventually evolve to other studies, namely multicriteria models for resource allocation.

Another important improvement for future work will be to streamline and optimize the participatory process, in order to reduce the time gap between meetings and the number of sessions necessary to create.

Furthermore, it is also important to improve and develop decision support tools, like the WisedOn software, allowing more functionalities, in order to expedite contact with decision-makers in future studies and works that will use this decision support system.

Besides, and as already mentioned by the decision-makers themselves as a final comment to this project, it would be interesting that, in an upcoming study (an extension of this project), the number of biomarkers under study was reduced, for a more in-depth analysis, where more evidence would be available. Nevertheless, this project was considered of extreme importance due to, not only to its originality in this area, but also since it served as a first filter to identify the most valuable emerging biomarkers.

Bibliography

[1] Canadian Medical Association, "The future of technology in health and health care: a primer," 2018.

[2] OECD, "New Health Technologies: Managing Access, Value and Sustainability, OECD Publishing, Paris.," tech. rep., 2017.

[3] A. G. Fraser, E. G. Butchart, P. Szymański, E. G. Caiani, S. Crosby, P. Kearney, and F. Van de Werf, "The need for transparency of clinical evidence for medical devices in Europe," 2018.

[4] MEDI-VALUE, "Developing HTA tools to consensualise MEDIcal devices' VALUE through multicriteria decision analysis." `http://medivalue.tecnico.ulisboa.pt/` (last accessed on 05/05/2021).

[5] MDR 745, "Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC and on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU," *Official Journal of the European Union*, pp. 2–175, 2017.

[6] Fortune Bussiness Inside, "Medical Devices market size, share and industry analysis by type." `https://www.fortunebusinessinsights.com/industry-reports/medical-devices-market-100085` (last accessed on 17/06/2020).

[7] Medicare Payment Advisor Commision, "An Overview of the Medical Device Industry," *Report to Congress: Medicare and the Health Care Delivery System*, pp. 207–242, 2017.

[8] MedTech Europe, "The European Medical Technology Industry – in figures 2019," 2019.

[9] Evaluate, "Evaluate MedTech: World Preview 2018, Outlook to 2024," tech. rep., 2018. `www.evaluate.com/MedtechWorldPreview2018` (last accessed on 17/06/2020).

[10] E. French-Mowat and J. Burnett, "How are medical devices regulated in the European Union?," *Journal of the Royal Society of Medicine*, pp. 22–28, 2012.

[11] World Health Organization, "Medical Devices: managing the mismatch," *World Health Organization*, pp. 1–36, 2010.

[12] European Comission, "Harmonised Standards." `https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en` (last accessed on 25/04/2021).

[13] Health Products Regulation Authority (HPRA), "Timeline and transition to the new Regulations." `http://www.hpra.ie/homepage/medical-devices/regulatory-information/new-eu-device-regulations/timeline-and-transition-to-the-new-regulations` (last accessed on 19/06/2020).

[14] Official Journal of the European Union, "REGULATION (EU) 2020/561 Medical Devices (amending Regulation (EU) 2017/745 on medical devices)," *Official Journal of the European Union*, pp. 1–5, 2020.

[15] World Health Organization, "WHO global model regulatory framework for medical devices including in vitro diagnostic medical devices," tech. rep., 2017.

[16] National Institute for Public Health and the Environment, "The impact of the new European IVD-classification rules on the notified body involvement," tech. rep., 2018.

[17] Biomarkers Definitions Working Group, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework," *Clinical Pharmacology and Therapeutics*, pp. 89–95, 2001.

[18] M. R. Bleavins, C. Carini, M. Jurima-Romet, and R. Rahbari, "Biomarkers," in *Pharmaceutical Sciencies Encyclopedia: Drug Discovery, Developmente, and Manufacturing*, p. 18, 2010.

[19] C. V. Gerlach, M. Derzi, S. K. Ramaiah, and V. S. Vaidya, "Industry Perspective on Biomarker Development and Qualification," *Clinical Pharmacology and Therapeutics*, pp. 27–31, 2018.

[20] National Library of Medicine, "PubMed." `https://pubmed.ncbi.nlm.nih.gov/` (last accessed on 23/03/2021).

[21] J. Beauchamp, K. Bedjama, T. Bracht, M. Breitner, J. Casado-Vela, R. Dégano, and F. Brambilla, *Biomarker Validation: Technological, Clinical and Commercial Aspects*. Wiley-VCH, 2015.

[22] Food and Drug Administration (FDA), "Approvals of FDA-Regulated products." `https://www.fda.gov/news-events/approvals-fda-regulated-products` (last accessed on 14/09/2020).

[23] FDA-NIH Biomarker Working Group, "BEST (Biomarkers, EndpointS, and other Tools) Resource," tech. rep., 2016.

[24] B. Chaffey and A. Silmon, "Biomarkers in personalized medicine: Discovery and delivery," *Biochemist*, pp. 43–47, 2016.

[25] Pharmaceutical Research and Manufacturers of America (PhaRMA), "Value of Personalized Medicine." `http://phrma-docs.phrma.org/sites/default/files/pdf/chart_pack-value_of_personalized_medicine.pdf` (last accessed on 30/08/2020).

[26] S. Mathur and J. Sutton, "Personalized medicine could transform healthcare (Review)," *Biomedical Reports*, pp. 3–5, 2017.

[27] K. K. Jain, "Role of Biomarkers in Personalized Medicine," in *Textbook of Personalized Medicine*, pp. 91–97, 2015.

[28] P. von Lode, V. Daussin, S. Raths, and S. Padurariu, "Best Practises and Pitfalls in Commercializing IVD-Applicable Biomarkers," tech. rep., 2020.

[29] World Health Organization (WHO), "HTA Definitions." `https://www.who.int/health-technology-assessment/about/Defining/en/` (last accessed on 31/05/2020).

[30] A. J. Culyer, "Health economics and health technology assessment," pp. 379–382, 2018.

[31] World Health Organization (WHO), *Global atlas of medical devices*. 2017. `http://www.who.int/medical_devices` (last accessed on 10/07/2020).

[32] European Science Foundation, "Implementation of Medical Reseach in Clinical Practice," 2011.

[33] L. K. Sampietro-Colom L, K. K. Cicchetti A, F. B. Pasternack I, W. C. Rosenmöller M, W. J. Kahveci R, and e. a. Kiivet RA, "The AdHopHTA handbook: a handbook of hospital-based Health Technology Assessment (HB-HTA)," tech. rep., AdHopHTA project, 2015.

[34] W. G. Goettsch, "Differences in Health Technology Assessment Recommendations Among European Jurisdictions: The Role of Practice Variations," *Value in Health*, pp. 10–16, 2020.

[35] M. P. Gagnon, "Hospital-Based Health Technology Assessment: Developments to Date," *PharmacoEconomics*, pp. 819–824, 2014.

[36] N. Martelli, P. Hansen, H. van den Brink, A. Boudard, A. L. Cordonnier, C. Devaux, J. Pineau, P. Prognon, and I. Borget, "Combining multi-criteria decision analysis and mini-health technology assessment: A funding decision-support tool for medical devices in a university hospital setting," *Journal of Biomedical Informatics*, pp. 201–208, 2016.

[37] N. Martelli, A. S. Lelong, P. Prognon, and J. Pineau, "Hospital-based health technology assessment for innovative medical devices in university hospitals and the role of hospital pharmacists: Learning from international experience," *International Journal of Technology Assessment in Health Care*, pp. 185–191, 2013.

[38] Danish Centre for Evaluation and Health Technology Assessment (DACEHTA), "INTRODUCTION TO MINI-HTA – a management and decision support tool for the hospital service," tech. rep., 2005.

[39] I. Masic, M. Miokovic, and B. Muhamedagic, "Evidence Based Medicine - New Approaches and Challenges," *Acta Informatica Medica*, p. 219, 2008.

[40] P. Cairney and K. Oliver, "Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy?," *Health Research Policy and Systems*, pp. 1–11, 2017.

[41] B. A. Teachman, D. A. G. Drabick, R. Hershenberg, D. Vivian, B. E. Wolfe, and M. R. Goldfried, "Bridging the Gap Between Clinical Research and Clinical Practice : Introduction to the Special Section," pp. 97–100, 2012.

[42] M. S. Pepe, H. Janes, C. I. Li, P. M. Bossuyt, Z. Feng, and J. Hilden, "Early-phase studies of biomarkers: What target sensitivity and specificity values might confer clinical utility?," *Clinical Chemistry*, pp. 737–742, 2016.

[43] X. H. Chen, S. Huang, and D. Kerr, "Biomarkers in clinical medicine," in *Molecular Epidemiology: Principles and Practises*, pp. 303–322, 2011.

[44] N. R. Cook, "Methods for evaluating novel biomarkers - A new paradigm," *International Journal of Clinical Practice*, pp. 1723–1727, 2010.

[45] M. G. Scott, "When do new biomarkers make economic sense?," *Scandinavian Journal of Clinical and Laboratory Investigation*, pp. 90–95, 2010.

[46] 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," *European Journal of Health Economics*, pp. 891–918, 2019.

[47] 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*, pp. 137–156, 2017.

[48] M. M. Goetghebeur, M. Wagner, H. Khoury, R. J. Levitt, L. J. Erickson, and D. Rindress, "Evidence and Value: Impact on Decision-Making - The EVIDEM framework and potential applications," *BMC Health Services Research*, pp. 1–16, 2008.

[49] London School of Economics, "Advance HTA - Advancing and strengthening the methodological tools and practices relating to the application and implementation of Health Technology Assessment (HTA)," p. 98, 2015.

[50] C. A. Bana E Costa, J. M. De Corte, and J. C. Vansnick, "Macbeth," *International Journal of Information Technology and Decision Making*, pp. 359–387, 2012.

[51] European Society for Medical Oncology (ESMO), "What is Breast Cancer?," pp. 1–55, 2018.

[52] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, "Breast cancer," *Nature Reviews*, 2019.

[53] R. Molina, V. Barak, A. Van Dalen, M. J. Duffy, R. Einarsson, M. Gion, H. Goike, R. Lamerz, M. Nap, G. Sölétormos, and P. Stieber, "Tumor markers in breast cancer - European group on tumor markers recommendations," 2005.

[54] A. Mehrgou and M. Akouchekian, "The importance of BRCA1 and BRCA2 genes mutations in breast cancer development," *Medical Journal of the Islamic Republic of Iran*, pp. 1–12, 2016.

[55] F. Cardoso, S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, I. T. Rubio, S. Zackrisson, and E. Senkus, "Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, pp. 1194–1220, 2019.

[56] World Health Organization, *WHO Classification of Tumors of the Breast*. 2012.

[57] National Comprehensive Cancer Network, "Breast Cancer - Invasive," tech. rep., 2018. `https://www.mdanderson.org/content/dam/mdanderson/documents/for-physicians/algorithms/cancer-treatment/ca-treatment-breast-invasive-web-algorithm.pdf` (last accessed on 27/10/2020).

[58] F. Cardoso, A. L. Vasconcelos, A. Ferreira, B. Sousa, and C. Loewenthal, *100 perguntas chave no cancro da mama*. 2017.

[59] S. Bertozzi, A. Londero, L. Seriau, R. Di Vora, C. Cedolini, and L. Mariuzzi, "Biomarkers in breast cancer," pp. 3–29, 2018.

[60] N. Iqbal and N. Iqbal, "Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications," *Molecular Biology International*, pp. 1–9, 2014.

[61] D. G. Hicks and S. Kulkarni, "HER2+ breast cancer: Review of biologic relevance and optimal use of diagnostic tools," *American Journal of Clinical Pathology*, pp. 263–273, 2008.

[62] I. De Paula Costa Monteiro, P. Madureira, A. De Vasconscelos, D. Humberto Pozza, and R. A. De Mello, "Targeting HER family in HER2-positive metastatic breast cancer: Potential biomarkers and novel targeted therapies," *Pharmacogenomics*, pp. 257–271, 2015.

[63] A. C. Wolff, M. Elizabeth Hale Hammond, K. H. Allison, B. E. Harvey, P. B. Mangu, J. M. Bartlett, M. Bilous, I. O. Ellis, P. Fitzgibbons, W. Hanna, R. B. Jenkins, M. F. Press, P. A. Spears, G. H. Vance, G. Viale, L. M. McShane, and M. Dowsett, "Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/ college of American pathologists clinical practice guideline focused update," *Journal of Clinical Oncology*, 2018.

[64] M. V. Dieci, F. Miglietta, G. Griguolo, and V. Guarneri, "Biomarkers for HER2-positive metastatic breast cancer: Beyond hormone receptors," *Cancer Treatment Reviews*, 2020.

[65] A. Madrid-Paredes, M. Canadas-Garre, A. Sánchez-pozo, and M. Á. Celleja-Hernández, "De novo resistance biomarkers to anti-HER2 therapies in HER2-positive breast cancer," *Pharmacogenomics*, 2015.

[66] M. G. Cesca, L. Vian, S. Cristóvão-Ferreira, N. Pondé, and E. de Azambuja, "HER2-positive advanced breast cancer treatment in 2020," *Cancer Treatment Reviews*, 2020.

[67] H. Saarikoski, D. Barton, J. Mustajoki, K. H., E. Gomez-Baggethun, and J. Langemeyer, "Multi-criteria decision analysis (MCDA) in ecosystem service valuation," *OpenNESS Ecosystem Service Reference Book*, pp. 1–5, 2015.

[68] L. D. Phillips and C. A. Bana E Costa, "Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing," *Annals of Operations Research*, no. 1, pp. 51–68, 2007.

[69] C. A. Bana e Costa, J. M. De Corte, and J.-C. Vansnick, "MACBETH," *LSE*, pp. 1689–1699, 2004.

[70] Decision Eyes, "Web-based Decision Support System for value assessment in Health Technology Assessment." `http://www.wised-on.com/` (last accessed on 02/05/2021).

[71] C. e Costa and E. Beinat, "Estruturação de Modelos de Análise Multicritério de Problemas de Decisão Pública," *Centro de Estudos de Gestão do Instituto Superior Técnico (CEG-IST)*, 2010.

[72] C. A. Bana e Costa, J. C. Lourenço, M. D. Oliveira, and J. C. Bana e Costa, "A Socio-technical Approach for Group Decision Support in Public Strategic Planning: The Pernambuco PPA Case," *Group Decision and Negotiation*, pp. 5–29, 2014.

[73] T. C. Rodrigues, "The MACBETH Approach to Health Value Measurement: Building a Population Health Index in Group Processes," *Procedia Technology*, 2014.

[74] Bana, C. e Costa, and E. Beinat, "Model-Structuring in Public Decision-Aiding," tech. rep., 2005.

[75] C. A. Bana e Costa, J. C. Lourenço, M. P. Chagas, and J. C. Bana e Costa, "Development of Reusable Bid Evaluation Models for the Portuguese Electric Transmission Company," *Decision Analysis*, pp. 22–42, 2008.

[76] C. A. Bana e Costa, J.-M. de Corte, and J.-C. Vansnick, "M-MACBETH (beta) - User's Guide," 2017.

[77] Elsevier, "Science Direct." `https://www.sciencedirect.com/` (last accessed on 23/03/2021).

[78] R. Nishimura, U. Toh, and M. Tanaka, "Role of HER2-Related Biomarkers (HER2, p95HER2, HER3, PTEN, and PIK3CA) in the Efficacy of Lapatinib plus Capecitabine in HER2-Positive Advanced Breast Cancer Refractory to Trastuzumab," *Oncology (Switzerland)*, pp. 51–61, 2017.

[79] B. Gyorffy, C. Hatzis, T. Sanft, E. Hofstatter, B. Aktas, and L. Pusztai, "Multigene prognostic tests in breast cancer: Past, present, future," *Breast Cancer Research*, pp. 1–7, 2015.

[80] T. Triulzi, G. V. Bianchi, and E. Tagliabue, "Predictive biomarkers in the treatment of HER2-positive breast cancer: An ongoing challenge," *Future Oncology*, pp. 1413–1428, 2016.

[81] B. d. A. Pultz, F. A. C. da Luz, P. R. de Faria, A. P. L. Oliveira, R. A. de Araújo, and M. J. B. Silva, "Far beyond the usual biomarkers in breast cancer: A review," *Journal of Cancer*, pp. 559–571, 2014.

[82] H. Fan, C. Li, Q. Xiang, L. Xu, Z. Zhang, Q. Liu, T. Zhang, Y. Zhou, X. Zhao, and Y. Cui, "PIK3CA mutations and their response to neoadjuvant treatment in early breast cancer: A systematic review and meta-analysis," *Thoracic Cancer*, pp. 571–579, 2018.

[83] G. Kutomi, T. Mizuguchi, F. Satomi, H. Maeda, H. Shima, Y. Kimura, and K. Hirata, "Current status of the prognostic molecular biomarkers in breast cancer: A systematic review," *Oncology Letters*, pp. 1491–1498, 2017.

[84] C. Solinas, L. Carbognin, P. De Silva, C. Criscitiello, and M. Lambertini, "Tumor-infiltrating lymphocytes in breast cancer according to tumor subtype: Current state of the art," *Breast*, pp. 142–150, 2017.

[85] S. E. Stanton and M. L. Disis, "Clinical significance of tumor-infiltrating lymphocytes in breast cancer," *Journal for ImmunoTherapy of Cancer*, pp. 1–7, 2016.

[86] L. A. Emens, "Breast cancer immunotherapy: Facts and hopes," *Clinical Cancer Research*, pp. 511–520, 2018.

[87] A. Ronchi, F. Pagliuca, F. Zito Marino, M. Accardo, I. Cozzolino, and R. Franco, "Current and potential immunohistochemical biomarkers for prognosis and therapeutic stratification of breast carcinoma," *Seminars in Cancer Biology*, pp. 1–9, 2020.

[88] M. Rohanizadegan, "Analysis of circulating tumor DNA in breast cancer as a diagnostic and prognostic biomarker," *Cancer Genetics*, pp. 159–168, 2018.

Appendices

Appendix A

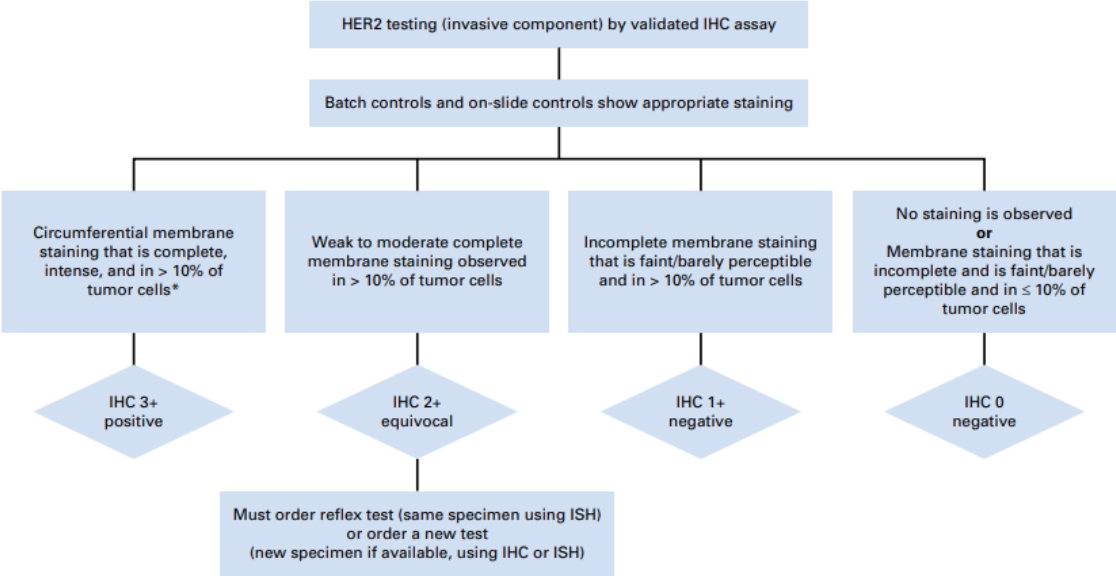# Algorithms for HER2 testing in breast cancer

Figure A.1: Algorithm of evaluation of HER2 expression by immunohistochemistry (IHC) assay of the breast cancer biopsy. [source: [63]]
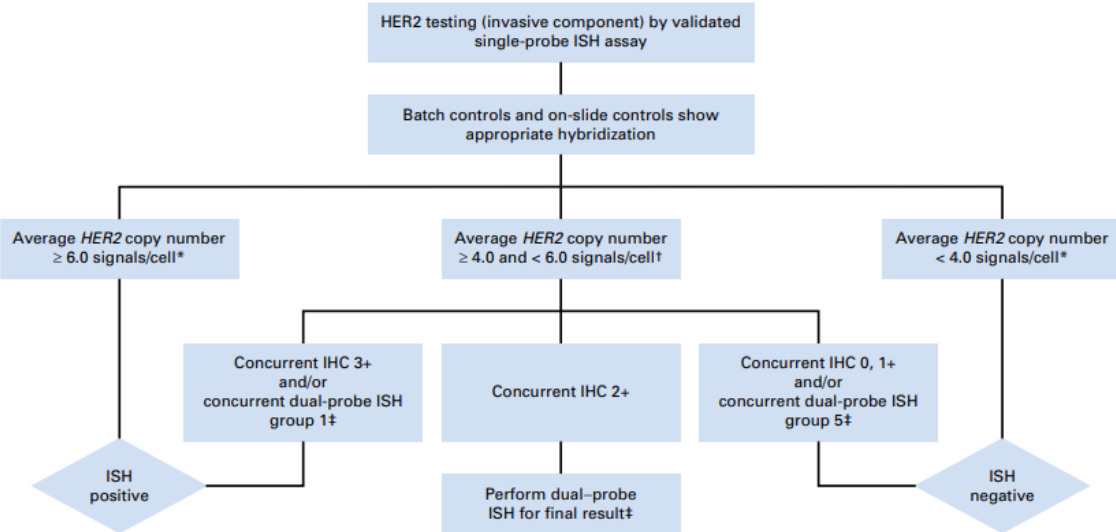

Figure A.2: Algorithm of evaluation of HER2 gene amplification by *in situ* hybridization (ISH) assay of the breast cancer biopsy using a single-signal (HER2 gene) assay (single-probe ISH). [source: [63]]
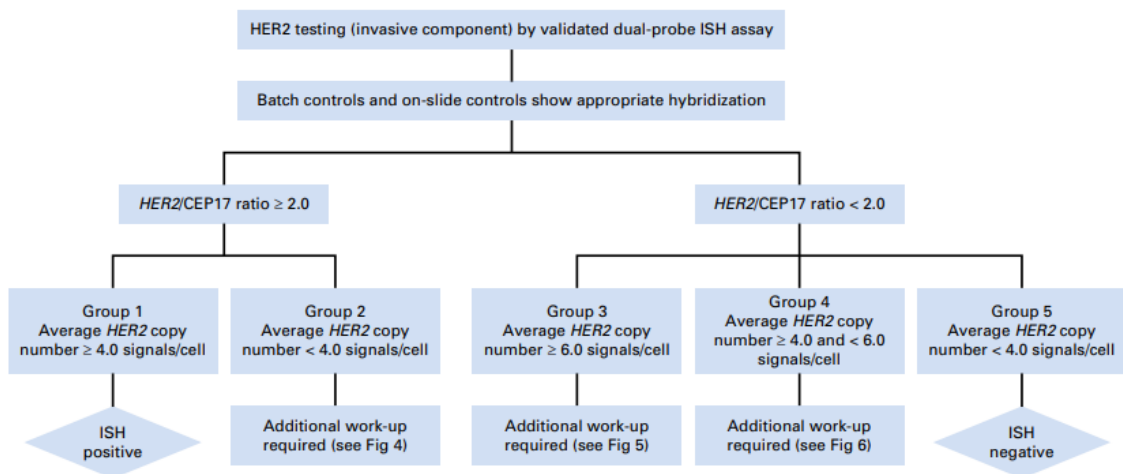
Figure A.3: Algorithm of evaluation of HER2 gene amplification by *in situ* hybridization (ISH) assay of the breast cancer biopsy using a a dual-signal (HER2 gene) assay (dual-probe ISH). Fig 4, Fig 5 and Fig 6 in the image corresponds to Figure A.4, Figure A.5 and Figure A.6, respectively. [source: [63]]
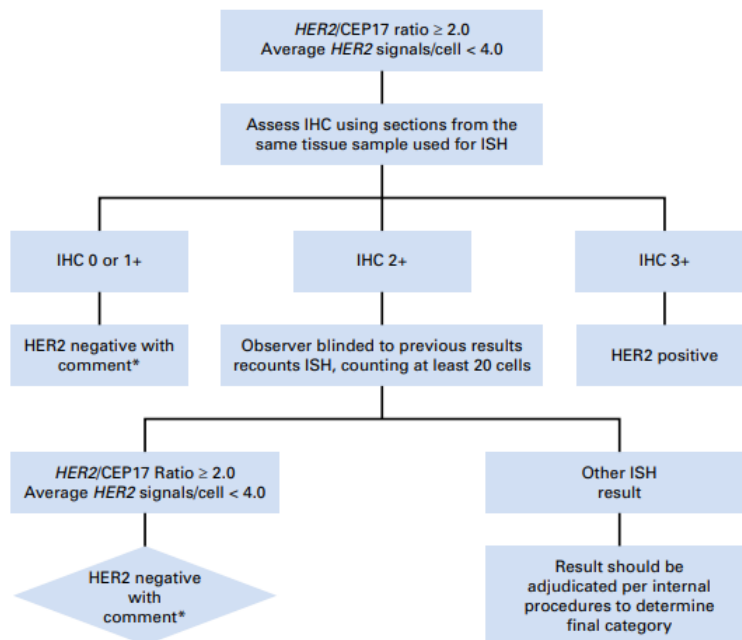

Figure A.4: Algorithm for additional work required from Group 2 (on Figure A.3). (*Comment: Evidence is limited on the efficacy of HER2-targeted therapy in the small subset of cases with HER2/CEP17 ratio $\geq$ 2.0 and an average HER2 copy number <4.0/cell. In the first generation of adjuvant trastuzumab trials, patients in this subgroup who were randomized to the trastuzumab arm did not appear to derive an improvement in disease free or overall survival, but there were too few such cases to draw definitive conclusions. IHC expression for HER2 should be used to complement ISH and define HER2 status. If IHC result is not 3+ positive, it is recommended that the specimen be considered HER2 negative because of the low HER2 copy number by ISH and lack of protein overexpression). [source: [63]]
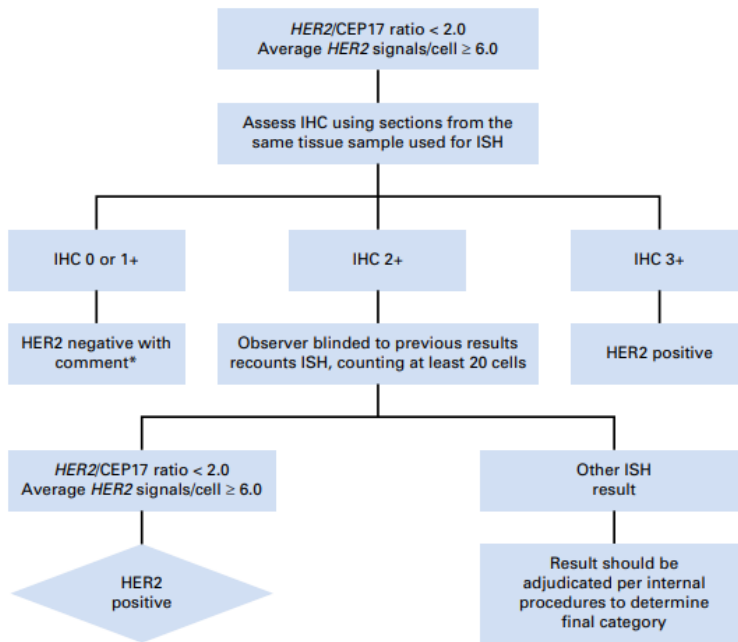
Figure A.5: Algorithm for additional work required from Group 3 (on Figure A.3) (*Comment: there are insufficient data on the efficacy of HER2-targeted therapy in cases with HER2 ratio <2.0 in the absence of protein over-expression because such patients were not eligible for the first generation of adjuvant trastuzumab clinical trials. When concurrent IHC results are negative (0,1+), it is recommended that the specimen be considered HER2 negative). [source: [63]]
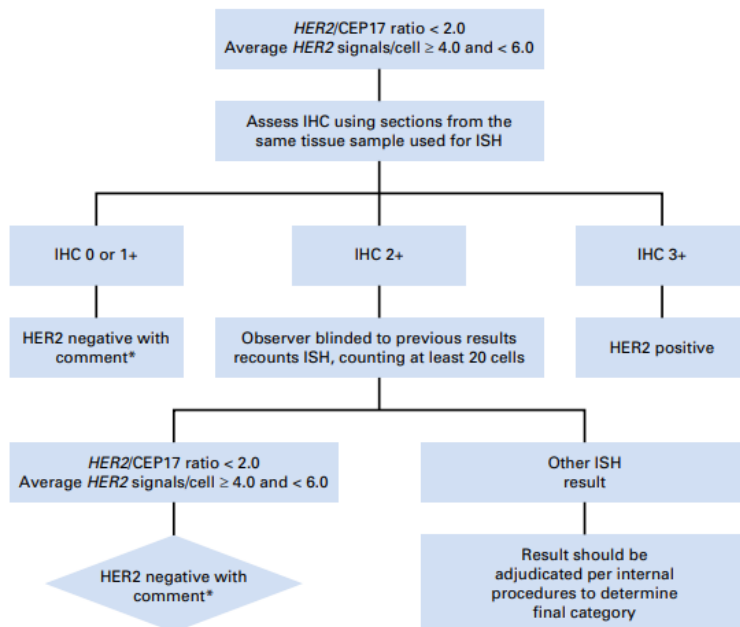
Figure A.6: Algorithm for additional work required from Group 4 (on Figure A.3). (*Comment: it is uncertain whether patients with ≥ 4.0 and <6.0 average HER2 signals/cell and HER2/CEP17 ratio <2.0 benefit from HER2 targeted therapy in the absence of protein overexpression (IHC 3+). If the specimen test result is close to the ISH ratio threshold for positive, there is a higher likelihood that repeat testing will result in different results by chance alone. Therefore, when IHC results are not 3+ positive, it is recommended that the sample be considered HER2 negative without additional testing on the same specimen). [source: [63]]

Appendix B

Matrices of judgement (for value scales)

	Very relevant, but it is not yet available in clinical settings	Indirect impact, is not yet available in clinical settings
Very relevant, it is already available in clinical settings	very strong	very strong
	Very relevant, but it is not yet available in clinical settings	strong

Figure B.1: Matrix of judgement for the criterion *Clinical relevance*. [source: WisedOn software]

	Approved at least by 1 local or national scientific group	Not yet recognize by any guideline
Approved at least by 1 international guideline	very strong	v. strong - extr
	Approved at least by 1 local or national scientific group	mod - strong

Figure B.2: Matrix of judgement for the criterion *Existence of clinical guidelines and quality of evidence*. [source: WisedOn software]

	External training, result is easy and simple to interpret	In-house training, result sometimes difficult to interpret	External training, result sometimes difficult to interpret
In-house training, result is easy and simple to interpret	strong	very strong	very strong
	External training, result is easy and simple to interpret	strong - v. strong	very strong
		In-house training, result sometimes difficult to interpret	strong

Figure B.3: Matrix of judgement for the criterion *Usability for the healthcare professional*. [source: WisedOn software]

	No impact on the workload, but with changes on the form of work	Impact on the form of work and workload
No impact on the form of work and workload	moderate	mod - strong
	No impact on the workload, but with changes on the form of work	strong

Figure B.4: Matrix of judgement for the criterion *Impact on the form of work and workload*. [source: WisedOn software]

	Quasi-unanimity (around 90%)	Qualified majority (around 75%)	Simple majority (around 51%)	Non-agreement (below 51%)
Unanimity (100%)	moderate	strong	extreme	extreme
	Quasi-unanimity (around 90%)	strong - v. strong	?	very strong
		Qualified majority (around 75%)	v. strong - extr	very strong
			Simple majority (around 51%)	weak

Figure B.5: Matrix of judgement for the criterion *Agreement of key-actors in the decision*. Note that the question mark (?) in a center cell of the matrix corresponds to a positive domain of the *good* reference level over the level of "Simple majority (around 51%)". [source: WisedOn software]

	Diagnostic and prognostic OR Prognostic and prediction of drug/therapy response	Diagnostic OR Prognostic OR Prediction of drug/therapy response
Diagnostic, prognostic and prediction of drug/therapy response	very strong	extreme
	Diagnostic and prognostic OR Prognostic and prediction of drug/therapy response	very strong

Figure B.6: Matrix of judgement for the criterion *Public health concern*. [source: WisedOn software]

Appendix C

Weighting matrix of judgement

	Clinical relevance (critical)	Public health concern (complementary)	Agreement of key-actors in the decision (fundamental)	Impact on the form of work and workload (fundamental)	Usability for the healthcare professional (complementary)	Neutral
Existance of clinical guidelines and quality of evidence (complementary)	v. strong - extr	very strong	very strong	very strong	v. strong - extr	very strong
	Clinical relevance (critical)	strong	very strong	strong - v. strong	very strong	strong - v. strong
		Public health concern (complementary)	very strong	v. strong - extr	v. strong - extr	very strong
			Agreement of key-actors in the decision (fundamental)	strong	strong - v. strong	strong
				Impact on the form of work and workload (fundamental)	moderate	mod - strong
					Usability for the healthcare professional (complementary)	moderate

Figure C.1: Weighting matrix of judgement. [source: WisedOn software]

89

Appendix D

Performance of the biomarkers

Biomarker	Criteria					
	Clinical relevance	Existence of clinical guidelines and quality of evidence	Usability for the healthcare professional	Impact on the form of work and workload	Agreement of key-actors in the decision	Public health concern
HER2 levels	L1	L1	L4	L2	L1	L3
HER2 mutations	L2	L1	L3	L1	L1	L2
HER2 heterogeneity	L1	L1	L3	L1	L1	L2
HER3	L1	L1	L3	L1	L1	L2
Intrinsic subtype (PAM50)	L3	L1	L3	L1	L1	L2
PTEN	L1	L1	L3	L1	L1	L1
PIK3CA mutations and PIK3 pathway inhibitors	L3	L1	L4	L2	L5	L1
TILs	L2	L2	L4	L2	L4	L2
PD-L1	L3	L3	L4	L2	L5	L1
FcγRs	L1	L1	L1	L1	L1	L1
Liquid Biopsy (ctDNA, ct-miRNA, CTC)	L2	L2	L3	L1	L4	L3

Figure D.1: Performance of the biomarkers in the different criteria.