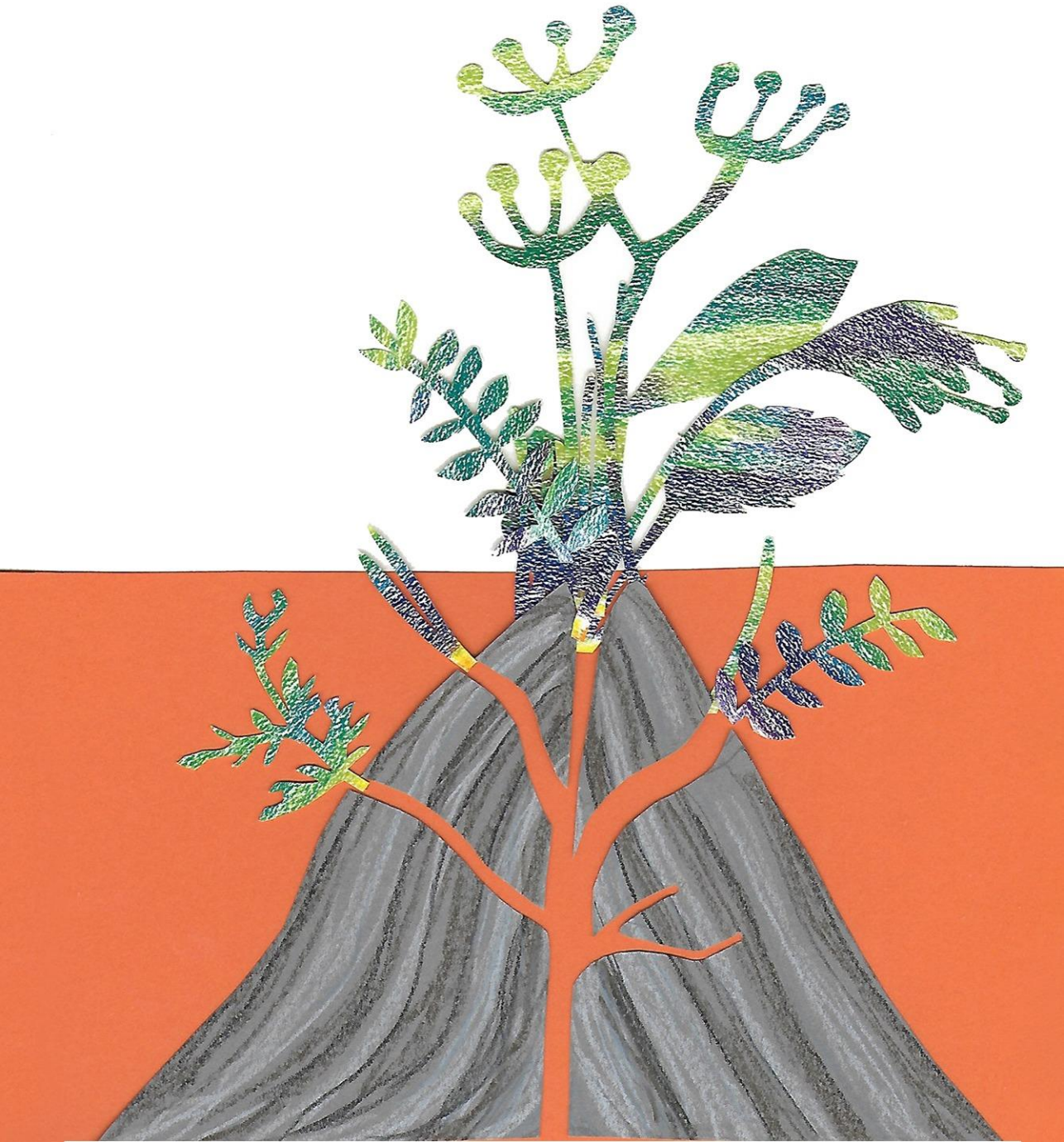
Economic evaluation of early warning systems for chronic disease management

The Case of Heart Failure

Fernando Albuquerque de Almeida



Economic Evaluation of Early Warning Systems for Chronic Disease Management

The Case of Heart Failure

Fernando Albuquerque de Almeida

ISBN 978-94-6423-388-9

Cover design Leonor Albuquerque de Almeida

Layout Fernando Albuquerque de Almeida

Printed by ProefschriftMaken // www.proefschriftmaken.nl

Economic Evaluation of Early Warning Systems for Chronic Disease Management

The Case of Heart Failure

Economische evaluatie van vroegtijdige signaleringssystemen voor chronische ziekten

De casus hartfalen

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the

rector magnificus

prof.dr. L.A. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Thursday, 28 October 2021 at 15:30 hrs

by

Fernando Guilherme França Pereira Albuquerque de Almeida

born in Lisbon, Portugal

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotor:	prof. dr. M.P.M.H. Rutten-van Mölken
Other members:	prof. dr. M.A. Joore
	prof. dr. J.J.M. Takkenberg
	prof. dr. W.B.F. Brouwer
Copromotors:	dr. M.J. Al
	dr. I. Corro Ramos

Table of contents

Chapter 1

General introduction

The burden of chronic disease

Chronic or noncommunicable diseases are broadly defined as conditions that last one year or more and that result of a combination of genetic, physiological, environmental, and behavioural factors, requiring ongoing medical attention or limiting activities of daily living (1, 2). The risk factors for chronic diseases can be categorised in modifiable behavioural risk factors, such as tobacco use, physical inactivity, unhealthy diet, and the harmful use of alcohol and metabolic risk factors like raised blood pressure, overweight/obesity, hyperglycaemia, and hyperlipidaemia (3).

An estimated 41 million people worldwide died of chronic diseases in 2016, representing 71% of all deaths. Four diseases were responsible for the vast majority of those deaths: cardiovascular diseases (17.9 million deaths), cancer (9.0 million deaths), chronic respiratory diseases (3.8 million deaths), and diabetes (1.6 million deaths) (2). In parallel with the high disease burden, the costs of managing chronic diseases take a huge toll on health care systems around the world, especially in highly developed countries. In the United States (US), circa 90% of the nation's $3.8 trillion in annual health care expenditure are attributable to patients with chronic and mental health conditions (4, 5). In the European Union (EU), every year, approximately 550,000 people of working age die prematurely from chronic disease, costing EU economies €115 billion (0.8% of GDP) in health care expenses (6). In 2016, 70% to 80% of the total healthcare costs in the EU – an estimated €700 billion – were spent on chronic diseases (7).

The *2030 Agenda for Sustainable Development* adopted by the United Nations (UN) recognised chronic disease management as one of the major challenges for improving health of the populations: unlike the advances against communicable diseases, the progress in the prevention and control of premature mortality from chronic diseases has lagged. Therefore, in order to achieve the World Health Organisation (WHO) *Sustainable Development Goals* targets by 2030, countries need comprehensive strategies to reduce death from chronic diseases more effectively (8). The current COVID-19 pandemic has reshaped public opinion on the importance of investing in health and healthcare, bringing into the public discussion the efforts that must be done in order to promote and safeguard public health. In the words of the Director-General of the WHO, Dr. Tedros Adhanom Ghebreyesus: *One of the key lessons from the COVID-19 pandemic is that we must invest in data and health information systems, as part of our overall public health capacity, before a crisis strikes. To emerge from this crisis stronger, we must be able to monitor progress with real-time, reliable and actionable data (8).*

Heart failure

Cardiovascular diseases lead as the disease group in chronic diseases with the highest burden in terms of disability-adjusted life-years (DALYs) across the globe (9). Heart failure (HF) is a type of cardiovascular disease described as a complex clinical syndrome that results from any structural or functional impairment of ventricular filling or ejection of blood (10, 11). HF is characterised by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) and signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral oedema) and it is caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress (12).

The main terminology used to describe HF is based on the measurement of the left ventricular ejection fraction (LVEF), which represents the percentage of blood that is ejected from the left ventricle of the heart with each beat. HF comprises a wide range of patients, from those with normal LVEF [typically considered as ≥50%; HF with preserved ejection fraction] to those with reduced LVEF [typically considered as <40%; HF with reduced ejection fraction]. Patients with an LVEF in the range of 40–49% are labelled as patients with mid-range ejection fraction (12). Each of those classes of patients according to their LVEF typically have different underlying aetiologies, demographics, co-morbidities, and response to therapy (13).

Another terminology related to the symptomatic severity of HF is the New York Heart Association (NYHA) functional classification (14). NYHA classes are used to describe the severity of symptoms and exercise intolerance. They provide useful and complementary information about the presence and severity of the disease, thus guiding HF treatment (see Table 1.1).

HF is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. An estimated 64.3 million people are living with heart failure across the globe (15). In developed countries, approximately 1% to 2% of the adult population live with HF, with its prevalence rising above 10% amongst the population older than 70 years of age; incidence is estimated to be between 5 and 10 per 1000 persons per year (16, 17). In addition, the absolute number of HF patients has been on the rise due to the aging population, global population growth, and improved survival after diagnosis (18, 19).

The costs related to heart failure in 2014 in the EU were estimated around €29 billion in one year (20). In the US, those costs were estimated at $30.7 billion in 2012, including the cost of health care services, medicines to treat heart failure, and productivity losses (21).

Table 1.1 – New York Heart Association functional classification based on severity of symptoms and physical activity: adapted from the American Heart Association website (22)

NYHA class	Patient symptoms
Class I	No limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea (shortness of breath).
Class II	Slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea.
Class III	Marked limitation of physical activity. Comfortable at rest. Less than ordinary activity causes fatigue, palpitation, or dyspnea.
Class IV	Unable to carry on any physical activity without discomfort. Symptoms of heart failure at rest. If any physical activity is undertaken, discomfort increases.

Decreased health-related quality of life (HRQoL) in HF patients has been reported in the literature and it has been associated with the symptomatology of the disease (23-26). More specifically, HRQoL has been shown to decrease as NYHA functional class worsens (27). HF entails a high hospitalisation burden, being responsible for about 1-2% of all hospital admissions (28) and the most common diagnosis in hospitalised patients older than 65 years of age (29, 30). After the first diagnosis, the average HF patient is hospitalised roughly once a year (31). Regarding mortality, since HF is a complex syndrome that can be viewed as the chronic stage of any underlying disease or condition leading to cardiac impairment, attributing an absolute number deaths due to HF can be challenging; HF-related mortality is often attributed to the most likely cause for death and not specifically to HF (32).

Numerous prognostic markers of death and/or HF hospitalisation have been identified in HF patients with HF (see Table 1.2). However, their direct clinical applicability may be limited and precise risk stratification in patients with HF remains challenging, in spite of the several multivariable prognostic risk scores developed for different HF populations (33-38). From a medical perspective, the goals of managing patients with HF consist in improving their clinical status, functional capacity, and quality of life, preventing hospital admissions, and reducing mortality (39-41).

Table 1.2 – Markers of worse prognosis in patients with heart failure: reproduced from 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure (12)

Demographic data	Older age, male sex, low socio-economic status.
Severity of heart failure	Advanced NYHA Class, longer HF duration, reduced peak oxygen consumption, high VE-VCO2 slope, Cheyne–Stoke ventilation, short 6-minute walking distance, reduced muscle strength, poor quality of life.

Clinical status	High resting heart rate, low blood pressure, clinical features of fluid overload (both pulmonary congestion and peripheral oedema, jugular venous dilatation, hepatomegaly), clinical features of peripheral hypoperfusion, body wasting, frailty.
Myocardial remodelling and severity of heart dysfunction	Low LVEF, LV dilation, severe diastolic LV dysfunction, high LV filling pressure, mitral regurgitation, aortic stenosis, LV hypertrophy, left atrial dilatation, RV dysfunction, pulmonary hypertension, dyssynchrony, vast area of hypo/akinesis, wide QRS complex, presumed inflammation or infiltration on CMR, inducible ischaemia and poor viability on imaging.
Biomarkers of neurohormonal activation	Low sodium, high natriuretic peptides, high plasma renin activity, high aldosterone and catecholamines, high endothelin-1, high adrenomedullin, high vasopressin.
Other biomarkers	Markers of renal function, inflammatory markers, cardiac stress markers, cardiac damage markers, metabolic markers, collagen markers, markers of organ damage/dysfunction.
Genetic testing	Certain mutations in inherited cardiomyopathies associated with high-risk of sudden cardiac death or rapid HF progression.
Cardiovascular co-morbidities	Atrial fibrillation, ventricular arrhythmia, non-revascularisable coronary artery disease, previous stroke/TIA, peripheral arterial disease.
Non-cardiovascular co-morbidities	Diabetes, anaemia, iron deficiency, COPD, renal failure, liver dysfunction, sleep apnoea, cognitive impairment, depression.
Non-adherence	Non-adherence with recommended HF treatment.
Clinical events	HF hospitalisation, aborted cardiac arrest, ICD shocks.

Abbreviations: CMR, cardiac magnetic resonance; COPD, chronic obstructive pulmonary disease; HF, heart failure; ICD, implantable cardioverter defibrillator; LV, left ventricular; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; QRS, Q, R, and S waves (combination of three of the graphical deflections); RV, right ventricular; TIA, transient ischaemic attack; VE-VCO_2, ventilatory equivalent ratio for carbon dioxide.

Chronic disease management and technology

Chronic disease management (CDM) consists in detecting, screening, and treating chronic diseases and providing access to palliative care for those in need (42). Managing chronic diseases is a major challenge for healthcare systems worldwide, which have been primarily designed to address acute episodic care rather than to provide organised care for people with long-term medical conditions (43). Chronic diseases often require long periods of supervision, observation, or care, which makes the basic features of primary care – including continuity, coordination, and comprehensiveness – the suitable setting for managing chronic conditions (44). In fact, evidence suggests that reorienting health policy and healthcare towards chronic

care systems – including strong primary care, which is more proactive in nature – generates better health outcomes at a lower cost than reactive systems relying mainly on secondary care (45, 46).

The Chronic Care Model is a well-established framework for CDM and for practice improvement. It includes key elements such as better integrated and coordinated care, collaboration across multidisciplinary teams of care providers, planned care with regular follow-up and review, and support for patient self-management (47). Despite evidence showing that that model leads to more effective care and improved patient outcomes for people with chronic diseases (48, 49), there is also evidence suggesting that many health care providers do not follow the recommended best practices effectively (50, 51).

Many barriers to the achievement of optimal CDM have been identified, including: (i) the complexity in the communication within the caring team, (ii) the complexity of designing meaningful care plans that are up-to-date, evidence-based, and personalised for the patient, (iii) the difficulty in keeping track of the actions of the caring team members and their responsibilities, (iv) the burden of regular review and follow-up, (v) the limited means of providing support for patient self-management, and (vi) the administrative overheads and bureaucracy associated with keeping clinical records (52-54).

Through tackling all the aforementioned barriers, healthcare practice supported by digital technologies (electronic processes and communications, the internet, and other information technologies) – usually branded under the broad name of eHealth, or mHealth if involving mobile devices – is expected to be a more effective way of approaching chronic disease management (55). Therefore, considering that existing health care delivery models seems to be outdated in order to effectively manage chronic disease – as evidenced by low adherence to quality and control indicators –, modifying health care delivery to include team-based care combined with patient-centred technologies seems to be a promising alternative (56) (see Figure 1.1).

The World Health Assembly in 2018 acknowledged the potential of digital technologies to play a major role in improving public health. Then, delegates agreed on a resolution on digital health that urges Member States to prioritize the development and greater use of digital technologies in health as a way of promoting *Universal Health Coverage* and advancing the *Sustainable Development Goals* (57).

The promise of technology in the treatment of chronic disease rests on two essential pillars: (i) providing a framework for patient engagement in changing modifiable behavioural risk factors and (ii) generating, collecting, treating, and analysing disease-related data that can be used for predicting important events related to the disease and for fine-tuning the treatment of patients.

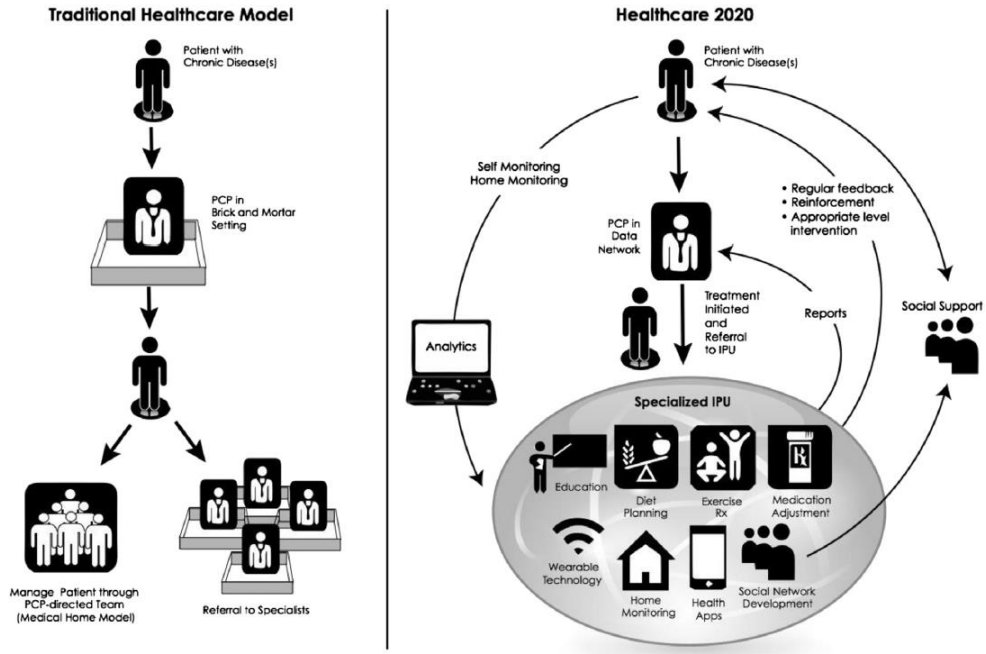


Figure 1.1 – Models of care delivery in chronic disease management: reproduced from Milani and Lavie (2015) (56)
IPU, integrated practice unit; PCP, primary care physician.

Early warning systems

Early warning systems (EWS) are timely surveillance systems that collect information on diseases in order to anticipate health deterioration and to trigger prompt clinical intervention, thereby improving prognosis and treatment outcomes (58). Generally speaking, EWS in health care consist of three main elements (59): (i) monitoring and collection of clinical data (e.g. vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a particular combination of signs and symptoms – that trigger an alarm and follow-up actions. Figure 1.2 presents a schematic representation of the operation of an early warning system.

In a first approach, we can think of EWS as simple algorithms/plans of action based on bedside observations for identifying patients at risk on general hospital wards (60). In this context, EWS are tools developed for recording physiological parameters (e.g. systolic blood pressure, heart rate, respiratory rate, urinary output, temperature, level of consciousness, etc.) and for using them to assess the level of risk for an undesirable

outcome through a composite score. This score is used as a mechanism for early intervention and for treatment to be initiated (61).

When seen in the light of the use of technology for chronic disease management, the implementation of EWS usually takes the form of any type of medical device for remote patient monitoring (RPM), which can be defined as a patient management approach that uses information and communication technologies to monitor and transmit physiological data related to patient health status between geographically separated individuals (62). RPM facilitates frequent or continuous assessment of disease signs and symptoms, which can be easily measured by patients, family or caregivers. It can lead to favourable health outcomes by improving patients' quality of life, by preventing the psychological and physical strain resulting from critical clinical events (e.g. hospital admissions, disease exacerbations, etc.), and by empowering patients and encourage them to take more responsibility for their own health (63, 64). RPM lays the foundation of a *hospital to home* framework, which may contribute to early discharge planning and to reduce hospital admissions as well as hospital stays. In this way, it present a promising patient management approach, especially for chronic diseases.

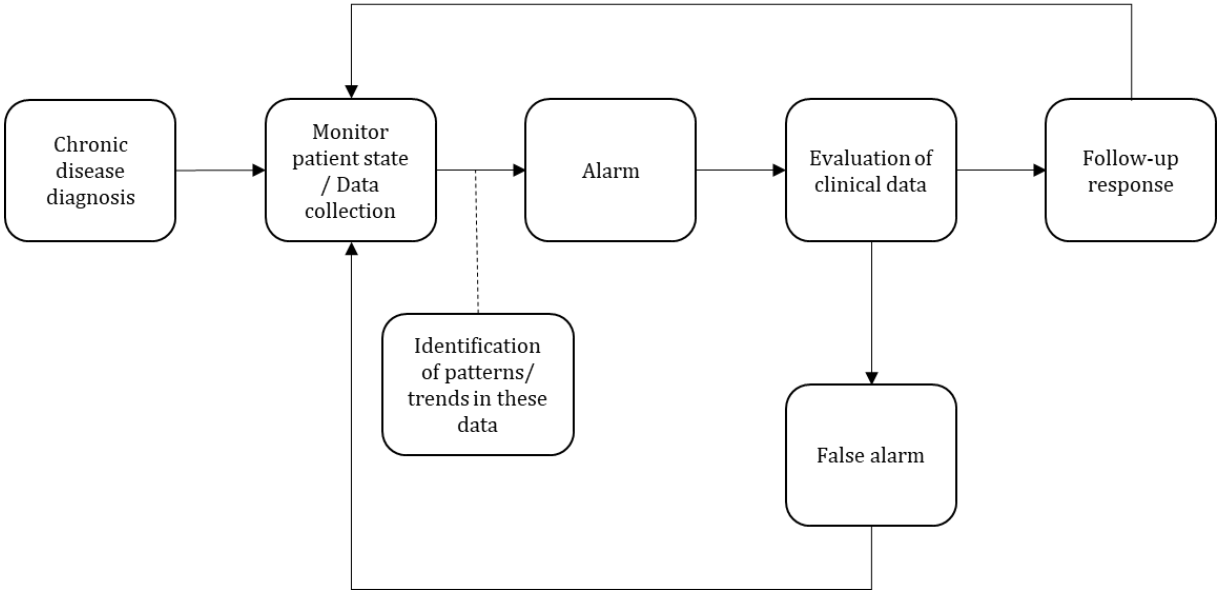


Figure 1.2 – Schematic representation of an early warning system

Nowadays computational potentialities allow for using more complex data analysis techniques, where the simple algorithms/plans of action mentioned above can be replaced by more complex methods. Diagnostic algorithms (DA) are mathematical relationships that use a wide range of data collected by EWS for calculating the likelihood of an event (e.g. hospitalisation or death). These algorithms are used for assisting medical personnel in their decision making process (65-68) by translating their output into clinical decision rules for clinical practice, for instance, by prioritising patients according to their likelihood of having an event or by raising an action-triggering alarm if the probability of having that event exceeds a pre-defined threshold (69).

Finally, it is also worth mentioning the huge potentialities of artificial intelligence and machine learning for continuously improving the DA prediction capabilities through the incorporation of big data collected by the EWS and/or other data sources, as well as for the constant fine-tuning of the follow-up actions resulting from a DA, based on the collected data on health outcomes (70, 71).

In conclusion, the future of EWS will be dictated by upcoming innovation in the field, the empowerment of patients – by involving them in their own health care –, and the acceptability of those interventions by health care professionals and patients. Given the unquestionable rise in the prevalence of chronic diseases, there will be an increase in the demand for innovative solutions. In reality, EWS are already playing a great role in shaping the future of health care and they are likely to see their importance grow sustainably.

Health technology assessment

Health technology assessment (HTA) is defined by the systematic evaluation of properties, effects, and/or impacts of health technology through a multidisciplinary process that assesses the social, economic, organisational and ethical issues of a health intervention or health technology with the main purpose informing a policy decision making (3).

HTA can be used in the scope of technology-related policies and decisions in a variety of ways (72, 73):

- Providing information on benefits and harms of new treatments compared to available treatment options;
- Determining reimbursement status;
- Supporting the price negotiation process;

- Helping clinicians and patients framing the use of an intervention for a particular clinical context;
- Informing health professional associations about the role of a technology in clinical protocols and/or practice guidelines;
- Supporting innovators in finding applications that are most likely to be (cost-)effective;
- Advising governments about the undertaking of public health programs (e.g., immunisation, screening, and environmental protection programs);
- Liaising with lawmakers and other political decision makers about policies on technological innovation, research and development, regulation, payment, and delivery of health care;
- Capacitating research agencies on evidence gaps and unmet medical needs;
- Prioritising innovations are steering innovation towards conditions with the highest unmet need;
- Supporting policy makers in planning resource capacity.

HTA has strong political support in most developed countries, where HTA agencies have been established for promoting rational choices informed by evidence on the allocation of limited resources for health care (74). The European Network for Health Technology Assessment (EUnetHTA), a joint effort aimed at creating an effective and sustainable network for HTA agencies across Europe (75), defines nine key domains for HTA activities (76): (i) health problem and current use of technology, (ii) description and technical characteristics of technology, (iii) safety, (iv) clinical effectiveness, (v) costs and economic evaluation, (vi) ethical analysis, (vii) organisational aspects, (viii) patient and social aspects, and (ix) legal aspects.

Economic evaluation

Economic evaluation (EE) is a topic of growing interest in the context of the assessment of health technologies, as policy makers have turned to evidence based decision making for supporting their political decisions.

For making those decisions, the more conventional welfarist economics aims at providing a coherent ethical framework for making meaningful statements about whether some states of the world are socially preferable to others (77). The neo-classical stream of welfarist economics is built on four key tenets (78): (i) the utility principle, where individuals rationally maximise their welfare by ordering options and choosing their preferred one; (ii) individual sovereignty, which determines that

individuals are the best judges of what contributes most to their utility; (iii) utility is derived only from the outcomes of behaviour and processes rather than the processes or intentions leading led to the outcomes; and (iv) welfarism, i.e. judging the goodness of states of affairs only by utility information. Under this line of decision making, interpersonal comparisons are normally disregarded and an overall social judgement is reached by using the Pareto principle, i.e. any increase of utility for one individual involves no utility loss for another (79).

Despite the welfarist approach being the most commonly used for making decisions in the health care space, an alternative extra-welfarist approach has been suggested, which differs from the first in four general principles: (i) by rejecting the tenet of welfarism that restricts the analysis to the individual utility – in health policy, other outcomes may include health or health gain, the distribution of health or health gain, and other measures like patient preferences or caregiver burden; (ii) by allowing the use of sources of valuation other than the affected individuals (e.g., experts, representatives of the general public, or authoritative decision makers); (iii) by permitting the weighting of relevant outcomes, which are often considered important as means of incorporating equity and other considerations, through a variety of ethical considerations including wealth, need, and desert; and (iv) by enabling the interpersonal comparison of relevant outcomes – although normally not in terms of individual utility, but rather in terms of capabilities and characteristics like health, handicap, ability to cope, schooling, ability to exercise discretion –, thus enabling movement beyond Paretian economics (80). In summary, the extra-welfarist approach proposes to offer the broadening of the evaluative space and the consequences that performing an evaluation may have in decision making.

In a practical manner, EE in health care can be defined as the comparison of two or more alternative healthcare interventions in terms of their costs and effects (81). EEs are labelled according to the way in which effects are measured, as made explicit in Table 1.3.

Cost-utility analysis (CUA) is the type of economic evaluation recommended in the guidelines of most jurisdictions (82). CUA incorporates HRQoL by measuring their results in quality-adjusted life years (QALYs) (83-85) – a measure of health in which the benefits in terms of length of life are adjusted to reflect the quality of life –, thus making comparisons across different diseases easier. However, despite many authors formally distinguishing between CUA and cost-effectiveness analysis (CEA), others consider CUA as a particular type of CEA (86), with the latter being often used as an umbrella term for both types of analyses, which will be the approach taken in this thesis.

Table 1.3 – Different economic evaluation studies: adapted from Drummond et al. (2015) (81)

	Denomination	Effects measured
Partial economic evaluation	Cost-minimisation analysis (CMA)	None: a cost-analysis of both alternatives is performed
	Cost-consequence analysis (CCA)	Health outcomes, adverse effects, etc. are listed and presented in a disaggregated tabular or graphical format
Full economic evaluation	Cost-effectiveness analysis (CEA)	Natural units (e.g. blood pressure, weight, life years
	Cost-utility analysis (CUA)	Utility (quality-adjusted life years, disability-adjusted life years)
	Cost-benefit analysis (CBA)	Monetary units

The cost-effectiveness results of a new intervention are usually expressed as an incremental cost-effectiveness ratio (ICER). The ICER represents the additional costs per an extra unit of effect gained with the new healthcare intervention (normally referred to as *intervention*) when compared to the current standard treatment (normally referred to as *comparator*). The ICER is achieved through the formula:

$$ICER = \frac{(Costs_{intervention} - Costs_{comparator})}{(Effects_{intervention} - Effects_{comparator})}.$$

If we take euros (€) as the currency used for measuring costs and the QALY as the unit used for measuring effects, the ICER will be presented in €/QALY. The ICER is then compared to a pre-defined cost-per-QALY threshold, which theoretically represents the opportunity costs of healthcare spending, given budget constraints, when seen from the supply side (87), or an estimate of the value that society places on a QALY, when seen from a demand side (88). Some authors argued that cost-effectiveness thresholds used for decision making are normally overestimated, as they are based on historical estimates, heuristics or judgements, and they should be replaced by empirical estimates of the supply side threshold, which could be considered more appropriate for judging the cost-effectiveness of new technologies when the aim is to maximise population health (89). However, beyond the academic discussion on the most appropriate way of defining the cost-effectiveness threshold, an intervention can be considered cost-effective when the ICER is lower than the pre-defined cost-effectiveness threshold. The cost-effectiveness plane (90), which plots the difference in costs in the Y-axis and the difference in effects in the X-axis between the intervention and the comparator, is an intuitive graph that is frequently used for presenting cost-effectiveness results (see Figure 1.3).

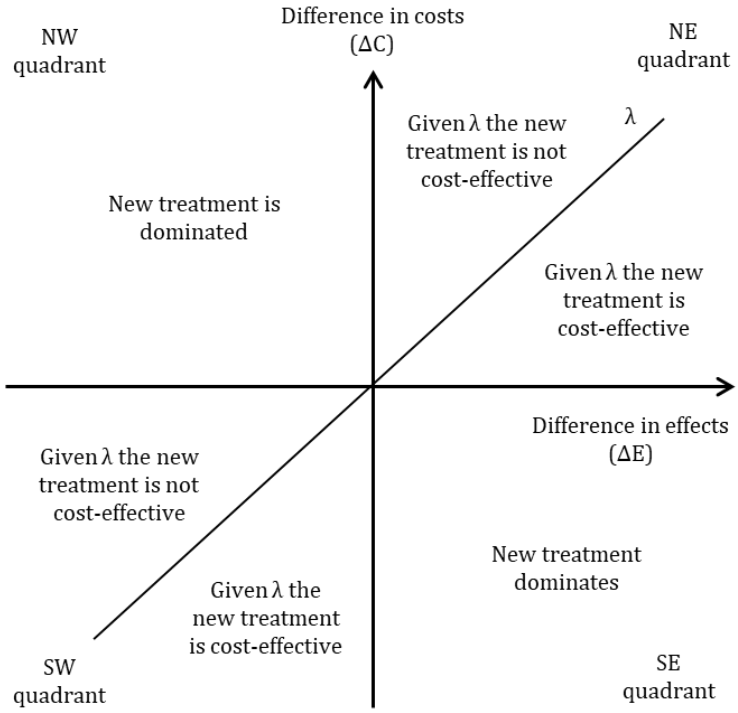


Figure 1.3 – Cost-effectiveness plane
λ, cost-per-QALY threshold; Δ, difference; NE, Northeast; NW, Northwest; SE, Southeast; SW, Southwest.

The uncertainty concerning the parameters used in economic evaluations is normally investigated through what we refer to as sensitivity analyses. Univariate or deterministic sensitivity analyses, when parameter values are individually changed and their impact on the ICER is assessed, or probabilistic sensitivity analyses, when a probabilistic distribution is attributed to each of the parameters used in the evaluation and the parameters are simultaneously varied a pre-defined number of times using Monte Carlo simulations (91). For each simulation, an ICER is calculated, thus allowing for the creation of confidence intervals around the ICER – using quantiles – and for the plotting of the simulations in the cost-effectiveness plane for a graphical visualisation of the impact of parameter uncertainty in the ICER.

An alternative to the use of the ICER is the incremental net monetary benefit (INMB), which can be calculated using the threshold (λ) as the value for the effects through the following expression: $INMB = \lambda \times \Delta E - \Delta C$, where ΔE and ΔC are the differences between effects and costs, respectively, between the intervention and the comparator. Thus, the ICER will be acceptable if: $ICER = \frac{\Delta C}{\Delta E} < \lambda$, for $\Delta E > 0$ and $ICER = \frac{\Delta C}{\Delta E} > \lambda$, for $\Delta E < 0$. In other words, the ICER is acceptable if the INMB is positive or, graphically, if the plotting of ΔE and ΔC in the cost-effectiveness plane is under the line for λ.

When comparing two interventions, the INMB concept can be applied to the simulations performed under the probabilistic sensitivity analysis. By calculating the percentage of simulations for which the INMB is positive at various cost-effectiveness thresholds, we can create a cost-effectiveness acceptability curve by plotting λ in the x-axis versus the probability of the intervention being cost effective in the y-axis (Figure 1.4) (92).

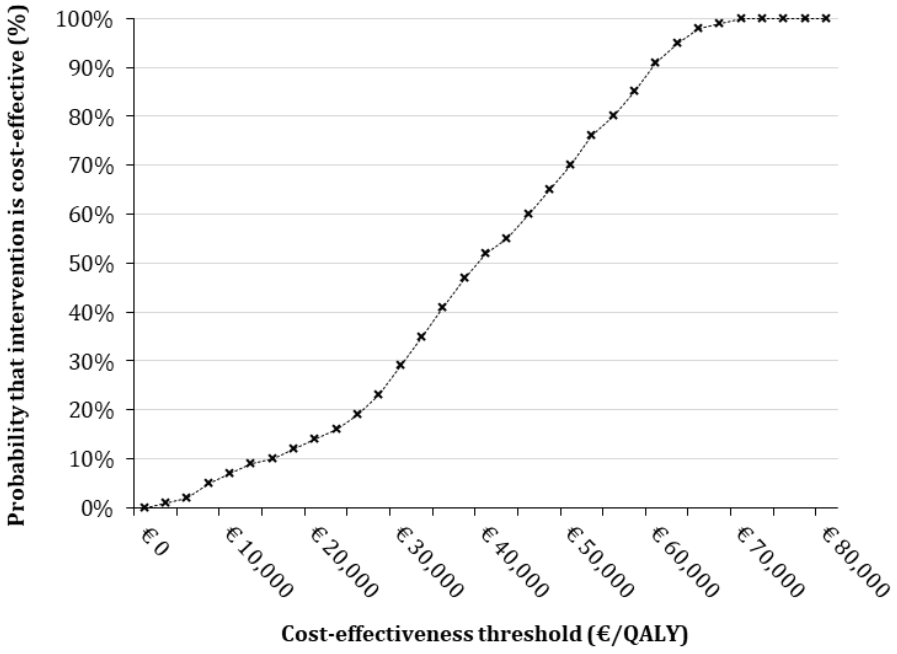


Figure 1.4 – Cost-effectiveness acceptability curve

Decision modelling

Because estimating the cost-effectiveness of an intervention in the health care field inevitably comprises the synthesis of information, the increasing use of economic evaluations for decision making in health care led to higher requirements in terms of analytic methodology. Firstly, researchers need to collect all the relevant information regarding the intervention(s) to be dealt with, normally related to factors including the epidemiology, natural history, costs, quality of life, and the implications of the analysed interventions on these parameters. Thereafter, they need to summarize all the available data, which is normally through the development of so called decision models.

Decision modelling can thus be defined as the set of analytical tools which researchers use to represent the complex reality in a more simplistic and comprehensible manner or by which experiments that are infeasible or impracticable are simulated (93).

The role of decision modelling in economic evaluation can be summarised in four different features (94): (i) synthesis of all relevant information in an analytical framework that reflects the possible prognoses and the disease pathways and their relationship with the interventions under evaluation; (ii) consideration of all relevant comparators, expanding from randomised control trials (RCTs), which are normally limited to head-to-head comparisons; (iii) using the appropriate time horizon for the context of decision making by extrapolating both costs and effects into the future; and (iv) addressing variability and uncertainty in a systematic and/or probabilistic manner.

In the context of economic evaluation, decision models use mathematical relationships to produce possible outcomes of a group of alternatives being evaluated. Based on a set of inputs defined by the user, a model is able to present a wide range of results that can be easily interpreted. It is precisely the generation of easily interpretable results – such as probability of an intervention being cost-effective at any given willingness-to-pay. This feature makes decision modelling extremely useful in practice, as the ultimate goal of economic evaluations is to inform decision makers so that they can make rational choices with regards to the allocation of resources. Therefore, the ability to decipher the methodology of decision modelling and to communicate results in a perceptible manner is the key to the success of decision modelling for the economic evaluation of health care interventions.

Decision modelling techniques are now completely established in HTA research, as health economic models represent strong analytical tools that empower decision making processes by contributing to the validity and generalisability of the results of economic evaluations (95).

Objectives

The main objective of this thesis is to study the methodology used in the economic evaluation of early warning systems for chronic disease management. More specifically, it focuses on the decision modelling methods used in this framework.

Decision models for assessing the cost-effectiveness of new healthcare interventions are normally disease-specific. However, given that EWS, regardless of their target disease, have in common that they are aimed at monitoring patients' health status through periodically measuring individual patient characteristics in order to anticipate health deterioration and to trigger prompt clinical intervention, it seems

worthwhile to explore the possibility to develop a more generic decision model for assessing their cost-effectiveness.

Following on the above, although this thesis intends to elaborate on the generic methods used in the economic evaluation of early warning systems for chronic disease management, heart failure is the disease and home telemonitoring the EWS intervention in focus, as these were the disease and intervention for which we had the necessary data for doing the proposed work in this thesis. The possible inferences for other diseases and early warning system interventions are discussed in the context of each chapter, where appropriate, and in the general discussion of the thesis.

Outline of the thesis

Besides the general introduction, this thesis has seven more chapters. **Chapter 2** consists of a systematic literature review describing the general and methodological characteristics of existing decision-analytic models for the economic evaluations of early warning systems for the management of chronic heart failure and a quality assessment of the methodological characteristics of those models. **Chapter 3** aims at determining the impact of nonfatal hospitalisations on the health-related quality of life of a cohort of patients previously diagnosed with heart failure by using their quality of life measurements before and after hospitalisation. **Chapter 4** describes a diagnostic algorithm for predicting the clinical deterioration in heart failure patients using a remote patient monitoring programme. **Chapter 5** presents the construction and validation of a discrete event simulation model that is able to model heart failure patients managed with usual care or an early warning system (with or without a diagnostic algorithm) and to account for the impact of individual patient characteristics in their health outcomes. **Chapter 6** uses the developed model for assessing the cost-effectiveness of a home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands. **Chapter 7** reviews and compares the legal framework in the United States and the European Union for the approval of medical devices and drugs and it compares the available information on clinical research and health technology assessment-supported recommendations in each of the considered jurisdictions for the health technologies under analysis. Finally, **Chapter 8** consists of the general discussion of the thesis, where its main findings are summarised, discussed, and interpreted in the context of the objectives of the thesis. Additionally, it issues recommendations for further research for healthcare policy.

Chapter 2

Early warning systems for the management of chronic heart failure: A systematic literature review of cost-effectiveness models

Fernando Albuquerque de Almeida, Maiwenn Al, Ron Koymans, Kadir Caliskan, Ankie Kerstens and Johan L. Severens

Expert review of pharmacoeconomics & outcomes research 18.2 (2018): 161-175.

Abstract

Introduction: Describing the general and methodological characteristics of decision-analytical models used in the economic evaluation of early warning systems for the management of chronic heart failure patients and performing a quality assessment of their methodological characteristics is expected to provide concise and useful insight to inform the future development of decision-analytical models in the field of heart failure management.

Areas covered: The literature on decision-analytical models for the economic evaluation of early warning systems for the management of chronic heart failure patients was systematically reviewed. Nine electronic databases were searched through the combination of synonyms for heart failure and sensitive filters for cost-effectiveness and early warning systems.

Expert commentary: The retrieved models show some variability with regards to their general study characteristics. Overall, they display satisfactory methodological quality, even though some points could be improved, namely on the consideration and discussion of any competing theories regarding model structure and disease progression, identification of key parameters and the use of expert opinion, and uncertainty analyses. A comprehensive definition of early warning systems and further research under this label should be pursued. To improve the transparency of economic evaluation publications, authors should make available detailed technical information regarding the published models.

Introduction

Heart failure (HF) is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. Throughout the Western world approximately 1–2% of the adult population has HF, with its prevalence rising to above 10% amongst persons ≥70 years of age; incidence is estimated to be between 5 and 10 per 1000 persons per year (16). In the United States alone, the total prevalence of heart failure was 5.1 million in 2010, with a reported incidence of 825,000 cases and an associated 57,757 deaths for that same year (96). Projections have shown that HF prevalence in the U.S. will increase 46% between 2012 and 2030, resulting in an increase of total medical costs from $20.9 billion in 2012 to $53.1 billion in 2030 (97). In the European Union, HF was responsible for total health care costs of €19.9 billion in 2009, circa 2% of the total health care expenditure for that year (98).

Heart failure is a condition characterized by typical symptoms (e.g., breathlessness, ankle swelling and fatigue) and signs (e.g., elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/ or elevated intracardiac pressures at rest or during stress. The current definition of HF relates to stages at which clinical symptoms are apparent. As such, before clinical symptoms become apparent, patients may have asymptomatic structural or functional cardiac abnormalities – such as systolic or diastolic left ventricular (LV) dysfunction –, which are precursors of HF (12). Severity of disease is usually measured using the New York Heart Association (NYHA) functional classification, which categorizes HF as mild (stages I and II), moderate (stage III), or severe (stage IV), based on the severity of the patients' symptoms (99). The goals of HF management are to relieve inherent signs and symptoms, prevent hospital admission, and improve survival (100). In this regard, not only mortality but also prevention of HF hospitalizations – which is now recognized as an important feature for patients and healthcare systems – have become the main outcomes of interest in clinical trials on HF (39).

Early warning systems are timely surveillance systems that collect information on diseases in order to anticipate health deterioration and trigger prompt clinical intervention, thereby improving prognosis and treatment outcomes (58). Generally speaking, early warning systems in health care consist of three main elements: (i) monitoring and collection of clinical data (e.g., vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a particular combination of

signs and symptoms – that trigger an alarm and follow-up actions. This entails more favourable health outcomes, with the concomitant improvement in the patients' quality of life, by buffering them from the psychological and physical strain resulting from hospital admissions. Additionally, via the monitoring of patients' vital signs and other important elements in an outpatient setting, early warning systems empower patients by encouraging them to take more responsibility for their own health (63, 64).

Considering that chronic heart failure management guidelines are targeted at relieving signs and symptoms, preventing hospital admission, and improving survival, it is expected that early warning system consist of a powerful tool in the management of the disease. Given that most of disease deterioration indicators are passible of being remotely monitored, early detection of these indicators enables better disease prognosis and treatment outcomes. There is already strong evidence in the literature that these systems can reduce both HF-related deaths and hospitalizations (101, 102), contrary to the conclusions of alternative studies that cast doubt on the effectiveness of outpatient management programmes (103). Additionally, avoiding hospitalizations using early warning systems may lead to substantial savings in costs. To gauge the sums involved, it should be recalled that each hospitalization for HF-related problems in the USA was reported to have an average cost of $18,000 in 2008 (104).

The ever increasing financial strains that plague health care systems confront their decision-makers with necessary choices about resource allocation. Understandably, cost-effectiveness became a colloquial term amongst decision-makers, who increasingly have been led to support their decisions on economic evaluations and evidence-based studies. The increased preponderance of these studies led to higher requirements in terms of analytic methodology, asking for the use of decision-analytic models – succinctly defined as sets of mathematical relationships that model the natural progression of disease and that, by simulating patient cohorts and disease pathways, allow for the estimation of clinical effects and their associated costs – that provide sound evidence for well-informed decision-making in the field of health technologies (105).

Describing the different approaches used in published models is expected to provide concise and useful insight to inform the future development of decision-analytical models in the field of heart failure management. Therefore, the objective of this study is to systematically review the literature on decision-analytical models used for the economic evaluation of early warning systems for the management of chronic heart failure patients, and to describe the general and methodological characteristics of those models.

Methods

Identification and selection of studies

A systematic literature review was performed to identify studies in which decision-analytical models were used for the assessment of the cost-effectiveness of early warning systems in the management of heart failure. The studies were selected according to the following inclusion criteria:

1. The study deals with a population ≥ 18 years old, of either sex or any ethnic group, who has been diagnosed with heart failure; both empirical (primary data) and theoretical (model) populations were considered.

2. The intervention under analysis is an early warning system as defined in the introductory section of this paper. No restrictions regarding to the comparator were imposed when selecting the studies for inclusion.

3. The study is an economic evaluation. Both cost-consequence analyses (106) or full economic evaluations (cost-effectiveness, cost-utility, or cost-benefit analyses) where there is a comparison between two or more alternatives and a simultaneous analysis of both costs and consequences (107) were considered.

4. The study reports patient and/or cost data, and uses a decision-analytical model based on such data. With regards to the term 'decision-analytical model', we used a modified definition from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Good Research Practices – Modeling Studies (108): “a logical mathematical framework that synthesizes evidence on health consequences and costs from many different sources, including data from clinical trials, observational studies, insurance claim databases, case registries, public health statistics, and preference surveys, whose purpose is to structure evidence on clinical and economic outcomes in a form that can help to inform decisions about clinical practices and health-care resource allocations.”

5. The study concerns an intervention which takes place in an outpatient setting.

6. The study is an English-language paper published in a peer-reviewed journal.

Taking into consideration the PICOS (population, intervention, comparator, outcomes, and study design) framework for study characteristics as defined in the PRISMA Statement (109), point 1 refers to the population, point 2 to the intervention and comparator, point 3 to the outcome, and points 4-6 to the study design.

Nine electronic databases (EMBASE, MEDLINE, CENTRAL, NHS Health Economic Evaluation Database, Health Technology Assessment database, Database of Abstracts of Reviews of Effects, Science Citation Index Expanded, PsycINFO, and Cumulative

Index to Nursing and Allied Health Literature) were searched up to December 2014 through the combination of synonyms for heart failure and sensitive filters for cost-effectiveness and early warning systems. The initial search strategy for EMBASE (searched via Embase.com) was developed using a combination of free text words, EMBASE-specific thesaurus terms, and a filter for excluding non-English articles and undesirable publications types (see Appendix 2.1 for the full search query). The remaining databases were searched after translating the search query using database-specific thesaurus terms and syntax. All identified citations from the electronic searches and other sources were imported into and managed using EndNote X6.

The selection of relevant articles resulted from a three-step process. Firstly, all titles were assessed by one reviewer (F.A.A.); studies were excluded if they explicitly failed to comply with one or more inclusion criteria. Secondly, two reviewers (F.A.A. and A.K.) independently screened abstracts of the remaining references. And thirdly, two reviewers (F.A.A. and M.A.) assessed the full-text articles for eligibility for data extraction. In steps two and three, studies were excluded if the reviewers considered that they did not meet all six inclusion criteria. Any divergences in those steps were resolved through discussion. If agreement could not be reached, a third reviewer (R.K.) made the final decision.

Analyses

Data extraction of general study characteristics

A table for data extraction was used for the abstraction of the general study characteristics. This table extracted model information concerning publication year, country, type of economic evaluation, objective, model type (Markov, decision tree, etc.), model structure, role of modelling (as described by Buxton et al. (93)), patient population, comparator, intervention, perspective of study, time horizon, cycle length, outcome(s) measured, uncertainty/sensitivity analyses, main results, and conclusion.

General characteristics of included studies were extracted independently by two reviewers (F.A.A. and J.S.). Disagreements were resolved by finding a consensus between the two reviewers. If consensus was not possible, a third reviewer (M.A.) was consulted.

Methodological characteristics assessment: Philips checklist

The methodological characteristics of included models were assessed using the checklist for the critical appraisal of decision-analytic models for health technology assessment developed by Philips et al. (110). Questions in the checklist were answered “yes” if the study paid objective attention to the item in question; “no” if the item was not fulfilled or insufficient information was provided to unequivocally score it as "yes"; “N/A” (not applicable) if the question was either not applicable to the study

or referred to a previous question scored with “no”; and “CT” (can’t tell) if the question could not be addressed given the information provided in the study.

The methodological characteristics assessment was undertaken independently by two reviewers (F.A.A. and J.S.). Dissimilarities in scoring were resolved by discussion. If agreement in a particular item was not possible the reasons for disagreement were disclosed to a third reviewer (M.A.), who attributed the final score for that item.

Results

Study retrieval

The electronic literature searches retrieved 6848 potentially relevant publications, from which 2083 were eliminated after deduplication. From the remaining 4765 references, 3636 were excluded based on title and further 1102 based on abstract, yielding 27 studies for full-text reading. The inclusion criteria further excluded 21 of these studies. However, the study by Burri et al. (111) was ultimately included – notwithstanding the fact that its target population was not restricted to patients diagnosed with heart failure but concerned all patients with cardiac implantable electric devices – on the belief that it contained information that would be valuable for answering our research question. Still, a formal distinction between this study and the remaining inclusions should be made (see Appendix 2.2 for the full list of excluded studies and reasons for exclusion).

In total we reviewed seven full-text publications reporting decision-analytical models (111-117). Figure 2.1 presents a flowchart describing the inclusion and exclusion process of relevant literature.

Analyses

General study characteristics

Information concerning the general and methodological characteristics of three of the reviewed models (113, 114, 116) could only be retrieved from additional publications in which fragments of these models were described. These publications were not technical reports but rather publications where the original model was described. Consequently, in order to fully evaluate the model reviewed in this study, four extra papers had to be included for assessment (118-121).

In Appendix 2.3 we provide a brief narrative description of the reviewed models, highlighting their general characteristics and giving an overview of the work that has been done in each of those studies. Table 2.1 provides an overview of the study characteristics of the reviewed models, showing existing differences and similarities between these characteristics.

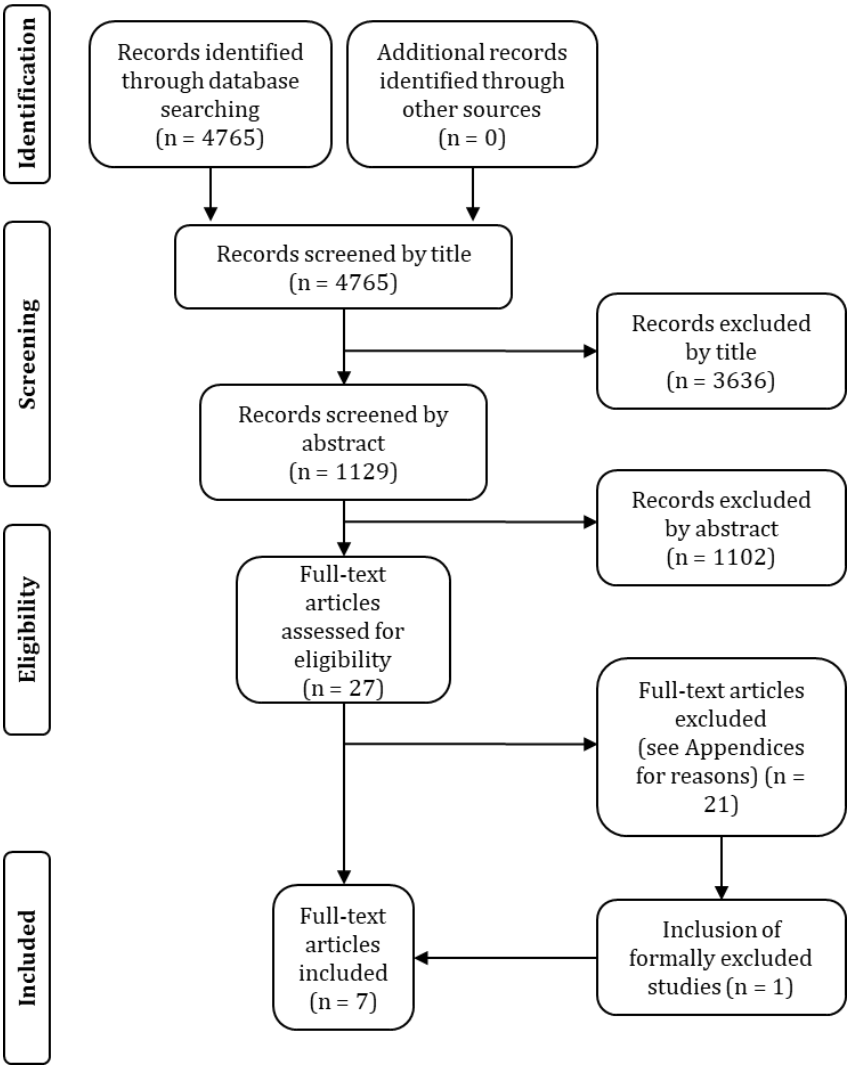


Figure 2.1 – Flowchart of the inclusion process of the search results

Except for Morimoto et al. (116), from 2004, all papers were published after 2009, within five years predating our review. All reviewed studies are cost-utility analyses, except the cost-consequence analysis by Burri et al. (111). The time horizon of the studies shows great variability, ranging from 1 year (112) to lifetime (113, 114). Only Burri et al. (111) does not report QALYs as a final outcome, and four articles (111, 112, 115, 117) report some sort of hospitalization measurement as an outcome of their studies. Five studies (111, 114-117) use Markov modelling – a type of model characterized by mutually exclusive health states which represent the possible consequences of the options under evaluation, and that reflect disease progression through stochastic transitions of patients between health states over discrete time periods (105), using two different types of Markov states: disease progression state (severity of disease, i.e. NYHA class) or health care provision state (hospitalization status). The remaining two studies (112, 113) model data using decision trees. Modelling plays three different roles in the reviewed studies: (i) informing decisions in the absence of hard data (111, 113, 116, 117), (ii) synthesizing head-to-head comparisons where relevant trials do not exist (112, 117), and (iii) extrapolating results of a single clinical trial over a longer timeframe (114, 115). With regards to the quantification of uncertainty, approaches are not consensual: only three studies (113, 115, 117) report probabilistic sensitivity analyses (PSA), while remaining models assess uncertainty merely by using deterministic sensitivity analyses; Pandor et al. (117) go one step further and include an expected value of perfect information (EVPI) analysis. The main results and conclusions of the reviewed articles are unanimous in their positive recommendations on the intervention(s) over the comparator.

Table 2.1 – General study characteristics

Study number, author, source	Burri et al. (111)	Klersy et al. (112)	Laramée et al. (113)	Miller et al. (114)	Moertl et al. (115)	Morimoto et al. (116)	Pandor et al. (117)
Publication year	2013	2011	2013	2009	2013	2004	2013
Country	UK	US, Italy, France, Germany, and UK	England and Wales	US	Austria and Canada	US	England and Wales
Type of economic evaluation	Cost-consequence analysis	Cost-effectiveness and cost-utility analyses	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis
Objective	Compare the long-term costs and consequences of daily home monitoring versus conventional follow-up in patients with cardiac implantable electric devices	Assess the cost-effectiveness and the cost-utility of remote patient monitoring versus with usual care	Assess the cost-effectiveness of three monitoring strategies for optimising medical therapy in chronic heart failure	Estimate the long-term cost-effectiveness of disease management programme in heart failure	Assess the cost-utility of NT-proBNP-guided, intensive patient management (BMC) versus multidisciplinary care or usual care	Assess the cost-utility of chronic heart failure management with or without B-type natriuretic peptide (BNP) measurement	Determine the clinical effectiveness and cost-effectiveness of home TM or STS strategies compared with UC for patients discharged from acute care after HF exacerbation
Model type (Markov, decision tree, etc.)	Markov model	Decision tree and budget impact model	Decision tree**	Markov model	Markov model	Markov model (two Markov models combined in decision tree)	Markov model
Model structure	Four health states: well, poststroke, post-ADHF, and dead	The decision tree in both arms considers two options: (i) the	Decision tree for intervention sub-divides in True HF and No HF, which	Four health states: NHYA I, NHYA II, NHYA III & IV, and dead	Five health states: 4 based on number of hospitalizations	Six health states: 5 according to number of hospitalizations	Four health states: alive at home, dead, HF-hospitalisation,

		patient is hospitalized for HF or (ii) the patient is not hospitalized for HF	respectively sub-divide in BNP above threshold (echo positive), BNP below threshold (HF missed), BNP above threshold (echo negative), and BNP below threshold (no echo)		(0, 1, 2, and ≥3) and death; after 18 months 9 health states, splitted according to beta blocker treatment	(0, 1, 2, 3, and ≥4) and death	and other hospitalization
Role of modelling* **	Informing decisions in the absence of hard data	Synthesizing head-to-head comparisons where relevant trials do not exist	Informing decisions in the absence of hard data	Extrapolating the results of a 18-month clinical trial	Extrapolating the results of a 18-month clinical trial	Informing decisions in the absence of hard data	Synthesizing head-to-head comparisons where relevant trials do not exist and informing decisions in the absence of hard data
Patient population	Patients who have undergone an implantable cardioverter defibrillators or cardiac resynchronizatio n therapy defibrillators implantation, and are managed in an outpatient setting*	Patients previously diagnosed with heart failure	Patients with CHF (CHF due to left ventricular systolic dysfunction and patients with CHF from any cause); subgroups ≤75 and >75 years	Male and female subjects, aged ≥18 years, with symptoms of CHF and documented systolic dysfunction	Patients discharged after heart failure hospitalization	Symptomatic CHF patients (NYHA classes II–IV) aged 35–85 after hospital admission because of CHF with leftventricular systolic ejection fraction < 40%	HF patients discharged from hospital within 28 days
Comparator	Conventional	Usual care: patient visit to a clinic	Usual care in the	Usual care	Usual care: patients were	CHF management without BNP	Usual care: standard post-

	follow-up	doctor's office, multidisciplinary outpatient clinic, or emergency department without additional phone calls from and to the patient	community		referred to their primary care physician with a detailed disease management plan	measurement	discharge multidisciplinary care without regular follow-up
Intervention	Daily home monitoring (BIOTRONIK Home Monitoring® system) where clinical and technical data transmitted automatically daily via the mobile phone network, plus instant automated alert transmission in case of a prespecified parameter deviation	Remote patient monitoring : (i) telephone monitoring approach with regular structured telephone contact between patients and healthcare providers; (ii) technology assisted monitoring approach relying on ICT, with data transfer via remote external monitors or CIEDs	(i) management guided by clinical assessment by a specialist; (ii) management guided by serial measurement of circulating NP concentration by a specialist	Telephonically administered DM programme consisting of a disease manager and a registered nurse with specialized cardiac training who provides patient education and medication management in conjunction with the primary care provider **	(i) nurse-led MC: visits by a specialized HF nurse, optional telephone support, and 2 prescheduled consultations with HF specialist; (ii) BMC with risk stratification performed upon NT-proBNP discharge levels	CHF management with BNP measurement	(i) home TM during office hours; (ii) home TM 24/7; (iii) human-to-human STS; and (iv) human-to-machine STS
Perspective of study	UK NHS	Health care payer	UK NHS and personal social services in England and Wales	Health care system	Health care payer	Health-care system perspective***	NHS in England and Wales
Time horizon	10 years	1 year	Lifetime	Lifetime	20 years	10 years**	30 years

Cycle length	1 year	N/A	1 year	6 months	1 month	3 months	1 month
Outcome(s) measured	12 health consequences and costs	Costs, # hospitalizations, length of stay for HF and for any cause, and QALYs	Costs, QALYs, and ICER	Costs, QALYs, and ICER	Costs, QALYs, and ICER	Costs, QALYs, and ICER	All-cause mortality, all-cause and HF-related admissions to hospital, costs, QALYs, and ICER
Uncertainty / Sensitivity analyses	Deterministic (univariate)	None	Probabilistic sensitivity analysis	Deterministic (univariate)	Deterministic and probabilistic sensitivity analyses	Deterministic (univariate and multivariate)	Probabilistic sensitivity analysis, EVPI analysis, and scenario analysis for costs
Main results	HM was cost neutral vs. CFU; HM reduced # patients with inappropriate shocks due to lead issues and atrial fibrillation; HM reduced the number of battery replacements and the number of in-office FU visits, while increasing unscheduled visits	The difference in costs between RPM and UC ranged from €300 to €1000, favouring RPM; QALY gain of 0.06 favouring RPM; length of stay once hospitalised similar for both groups	NP was the most cost-effective option in patients with CHF due to LVSD (ICER £3,304 vs. CA) and from any cause for all patients (ICER £14,694 vs. CA) and subgroup <75 years (ICER £2,517 vs. UC); NP was dominated by other strategies in subgroup >75 years with CHF from any cause, where CA was cost-effective vs. UC (ICER £11,508)	DM programme increased overall costs per patient by $4,850 whilst increasing QALYs by 0.111, leading to an ICER of $43,650; results were robust for changes in mortality rates, costs of care due to aging, utility values and the targeted population; assuming all programme costs to be variable led to an ICER of $129,738	MC vs. UC with ICERs of €3,746 for Austria and $5,554 for Canada; BMC strategy dominant over both MC and UC; cost-effectiveness acceptability curves show highest likelihood of BMC being the most cost-effective alternative at different thresholds	Additional QALYs (0.57 vs. 0.55) and lower costs ($9,577 vs. $10,131) for the BNP group when compared to non-BNP group; BNP group dominant over comparator	24/7 TM was not evaluated due to lack of data; TM during office hours resulted in £11,873 per QALY vs. UC, whereas STS-HH had an ICER of £228,035 per QALY when compared with TM during office hours; STS-HM was dominated by other alternatives

Conclusion	The model establishes HM as an economically viable technology when applied to the UK NHS	The cost-effectiveness data coupled with the demonstrated clinical efficacy of RPM vs. UC should encourage the acceptance of RPM amongst clinicians and consideration by third-party payers	NP is the most cost-effective strategy for CHF due to LVSD and from any cause, except in the subgroup of patients >75 years with CHF from any cause	Estimation of the clinical benefits and financial burden of DM can be enhanced by model-based analyses; results suggest that DM of HF patients can be cost-effective over the long term	NT-proBNP guided, intensified HF specialist patient management in addition to multidisciplinary care is dominant over multidisciplinary care alone and usual care	Introduction of BNP measurement in heart failure management may be cost-effective	Despite wide variation in UC and RM, cost-effectiveness analyses suggest that home TM during office hours was the optimal strategy in most costing scenarios

Abbreviations: #, number; ADHF, acute decompensated heart failure; BMC, NT-proBNP-guided, intensive patient management; BNP, brain natriuretic peptides; CA, clinical assessment; CHF, chronic heart failure; CIED, cardiac implantable electric devices; CFU, conventional follow-up; DM, disease management; EVPI, expected value of perfect information; HF, heart failure; HM, home monitoring; ICD, implantable cardioverter defibrillators; ICER, incremental cost-effectiveness ratio; ICT, information communication technology; LVSD, left ventricular systolic dysfunction; MC, multidisciplinary care; N/A, not applicable; NHS, National Health Service; NP, natriuretic peptide; NYHA, New York Heart Association; QALY, quality-adjusted life year; RPM, remote patient monitoring; STS, structured telephone support; STS-HH, human-to-human STS; STS-HM, human-to-machine STS; TM, telemonitoring; UC, usual care; UK, United Kingdom; US, United States

* Reason for formal exclusion

**Information to score the particular item was only retrieved from the additional publication in which part of the model was described

***Inferred due to reported data

****Based on Buxton et al. (93)

Methodological characteristics: quality assessment using the Philips checklist

Table 2.2 displays the results of the quality assessment using the checklist for the critical appraisal of decision-analytic models (110). Items in several topics were not fulfilled or insufficient information was available in the article in order to accurately assess them. Bearing in mind this review's objective of providing a useful insight for the future development of decision-analytical models in HF management, some missing information was considered critical. A few items from the rationale for structure and structural assumptions were not addressed sufficiently in most of the studies, with special notice to the consideration and discussion of any competing theories regarding model structure and disease progression [only one out the seven (Pandor et al. (117)) address this issue appropriately]. Furthermore, several items of the data identification section were not addressed in proper detail, especially concerning the identification of key parameters (only two studies (111, 114) justified the process of selecting key parameters and used systematic methods to identify the most appropriate data) and the use of expert opinion (none of the studies described and justified the use of expert opinion inputs in their work). Finally, the four types of uncertainty (methodological, structural, heterogeneity, and parameter) were not discussed or addressed in sufficient detail in all of the reviewed studies.

Table 2.2 – Scores on the quality assessment framework for decision-analytic models

Study number, author, source			1. Burri et al. (111)	2. Klersy et al. (112)	3. Laramée et al. (113)	4: Miller et al. (114)	5. Moertl et al. (115)	6. Morimoto et al. (116)	7. Pandor et al. (117)
Structure (S)									
S1: Statement of decision problem/objective	1	Is there a clear statement of the decision problem?	Y	Y	Y	Y	Y	Y	Y
	2	Is the objective of the evaluation and model specified and consistent with the stated decision problem?	Y	Y	Y	Y	Y	Y	Y
	3	Is the primary decision maker specified?	N	N	N	N	N	N	Y
S2: Statement of scope/perspective	1	Is the perspective of the model stated clearly?	Y	Y	Y	Y	Y	N	Y
	2	Are the model inputs consistent with the stated perspective?	N	N	N	Y	Y	N/A	Y
	3*	Has the scope of the model been stated and justified?	Y	N	N	Y	Y	N	Y
	4	Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	N	N	N	Y	Y	N/A	Y
S3: Rationale for structure	1	Has the evidence regarding the model structure been described?	Y	Y	Y	Y	Y	Y	Y
	2	Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	Y	N	N	N	N	N	N

	3	Have any competing theories regarding model structure been considered?	N	N	N	N	N	N	Y
	4	Are the sources of data used to develop the structure of the model specified?	Y	Y	Y	Y	Y	Y	Y
	5	Are the causal relationships described by the model structure justified appropriately?	Y	Y	Y	Y	N	N	Y
S4: Structural assumptions	1	Are the structural assumptions transparent and justified?	Y	N	N	Y	Y	Y	Y
	2	Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	Y	N	N	N	N	N	N
S5: Strategies/comparators	1	Is there a clear definition of the options under evaluation?	N	Y**	Y**	Y**	Y	Y	Y
	2	Have all feasible and practical options been evaluated?	CT	CT	CT	CT	CT	CT	CT
	3	Is there justification for the exclusion of feasible options?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
S6: Model type	1	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	N	Y	Y	Y	Y	Y	Y
S7: Time horizon	1	Is the time horizon of the model sufficient to reflect all important differences between options?	N	N	N	Y	Y	N	Y
	2	Is the time horizon of the model, and the duration of treatment and treatment effect described and justified?	Y	Y	Y	Y	Y	Y**	Y

	3	Has a lifetime horizon been used?	N	N	N	Y	N	N	N
	4	If not, has a shorter time horizon been justified?	Y	Y	Y	N/A	N	N	Y
S8: Disease states/pathways	1	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	Y	Y	Y	Y	N	N	N
S9: Cycle length	1	Is the cycle length defined and justified in terms of the natural history of disease?	N	N/A	N/A	Y	N	Y	Y
Data (D)									
D1: Data identification	1	Are the data identification methods transparent and appropriate given the objectives of the model?	Y**	Y	Y	Y	N	Y	Y
	2	Where choices have been made between data sources, are these justified appropriately?	Y**	Y	Y	Y	Y	N/A	N/A
	3	Has particular attention been paid to identifying data for the important parameters in the model?	N	N	N	N	N	N	Y
	4	Has the process of selecting key parameters been justified and systematic methods used to identify the most appropriate data?	Y	N	N	Y	N	N	N
	5	Has the quality of the data been assessed appropriately?	CT	Y	Y	N	N	Y	Y
	6	Where expert opinion has been used, are the methods described and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N

D2: Pre-model data analysis	1	Are the pre-model data analysis methodology based on justifiable statistical and epidemiological techniques?	CT	Y	Y	Y	CT	CT	Y
D2a: baseline data	1	Is the choice of baseline data described and justified?	Y**	Y	Y	Y	Y	Y	Y
	2	Are transition probabilities calculated appropriately?	CT	N/A	N/A	Y	CT	CT	Y
	3	Has a half cycle correction been applied to both cost and outcome?	N	N/A	N/A	N	N	N	Y
	4	If not, has this omission been justified?	N	N/A	N/A	N	N	N	N/A
D2b: treatment effects and diagnostic accuracy	1	If relative diagnostic accuracy have been derived from trial data, have they been synthesised using appropriate techniques?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	2	Have the methods and assumptions used to extrapolate diagnostic accuracy to final outcomes been documented and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	3	Have alternative assumptions been explored through sensitivity analysis?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	4	Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	N	N/A	N/A	N	N	Y	Y
	5	Have alternative assumptions been explored through sensitivity analysis?	N	N/A	N/A	N	N	Y	Y

D2c: quality-of-life weights (utilities)	1	Are the utilities incorporated into the model appropriate?	N/A	Y	Y	Y	Y	Y	Y
	2	Is the source for the utility weights referenced?	N/A	Y	Y	Y	Y	Y	Y
	3	Are the methods of derivation for the utility weights justified?	N/A	N/A	N/A	Y	N	Y**	N/A
D3: Data incorporation	1	Have all data incorporated into the model been described and referenced in sufficient detail?	Y	Y	Y	Y	Y	Y	Y
	2	Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	CT	N/A	N/A	CT	CT	CT	CT
	3	Is the process of data incorporation transparent?	Y	Y	Y	Y	Y	Y	Y
	4	If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	N/A	Y	Y	N	N	N	Y
	5	If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	N/A	N/A	N/A	N	N/A	N/A	Y
D4: Assessment of uncertainty	1	Have the four principal types of uncertainty been addressed?	N	N	N	N	N	N	N
	2	If not, has the omission of particular forms of uncertainty been justified?	N	N	N	N	N	N	N
D4a: methodological	1	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	N	N	N	N	N	N	N

D4b: structural	1	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	N	N	N	Y	N	N	N
D4c: heterogeneity	1	Has heterogeneity been dealt with by running the model separately for different subgroups?	N	N	N	Y	N	N	N
D4d: parameter	1	Are the methods of assessment of parameter uncertainty appropriate?	N	Y	Y	Y	Y	Y	Y
	2	Has probabilistic sensitivity analysis been done, if not, has this been justified?	N	N	N	N	Y	N	Y
	3	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	N	Y	Y	Y	N	N	N/A
Consistency (C)									
C1: Internal consistency	1	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	Y	N	N	Y	Y	N	N
C2: External consistency	1	Are the conclusions valid given the data presented?	Y	Y	Y	Y	Y	Y	Y
	2	Are any counterintuitive results from the model explained and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	3	If the model has been calibrated against independent data, have any differences been explained and justified?	N	N/A	N/A	N	N	N	N
	4	Have the results of the model been compared with those of previous models and any differences in results explained?	Y	N	N	Y	Y	N	Y

Abbreviations: Y, both reviewers agreed that the study paid sufficient attention to an item; N, the item was not fulfilled or insufficient information was provided to unequivocally score it as "yes"; N/A (not applicable), if the question was either not applicable to the study or referred to a previous question scored with “no”; "CT" (can't tell) if the question could not be addressed given the information provided in the study.

*According to the framework, the model scope should include the perspective, involved technologies, population, setting, and time horizon at the outset of the study.

**Information to score the particular item was only retrieved from the additional publication in which part of the model was described.

Discussion

Seven studies were retrieved while systematically reviewing the literature presenting decision-analytical models utilized for the economic evaluation of early warning systems for the management of chronic heart failure patient. The quality of these models was assessed through the checklist developed by Philips et al. (110) for the critical appraisal of decision-analytic models in health technology assessment. Overall, the models show satisfactory methodological quality, even though some points for improvement may be glimpsed.

Bearing in mind that the comparison of cost-effectiveness outcomes was not the main focus of the present study, it is worth noting that the wide range of study characteristics and methodologies displayed in the reviewed articles may have a major impact on the cost-effectiveness estimates, thereby impairing the soundness of the comparisons. Furthermore, the wide variety of model approaches, leaves room for standardization in a research field that is far from being consensual. Guidelines for good modelling practice (108) and key principles in health technology assessment (122) may prove to be highly helpful for decision-analytic modellers. Models tend to reflect the diverse decision support prevailing in different countries and their changes through time. This brings up the issue of transferability, i.e. when data from an economic evaluation done in one geographic area is transferred to another location or to another moment in time (123).

During the selection process we found many studies that did not perform a full economic evaluation but rather focused only on the effectiveness side of the cost-effectiveness equation (124-126). In the study by Pandor et al. (117) the authors performed a systematic review of the literature and meta-analysis to evaluate the clinical effectiveness of home telemonitoring or structured telephone support strategies – which both fall under the label of early warning systems – when compared with usual care for adults who have been recently discharged (within 28 days) from an acute care setting after a recent exacerbation of HF. To our knowledge the study by Pandor et al. (117) is the only one which partially overlaps with our review. Following the systematic review to evaluate the clinical effectiveness, the authors added an economic filter to review cost-effectiveness evidence on the same interventions and isolated two studies (112, 114) which were also reviewed in this paper. An assessment of the methodological quality of each study was performed using a combination of key components of the Drummond and Jefferson checklist for economic evaluations (107, 127), and the Eddy checklist for mathematical models used in technology assessments (128). Finally the authors built an economic model to evaluate the cost-effectiveness of several strategies for remote monitoring compared with usual care for patients recently discharged with heart failure. Comparing our

work with the one by Pandor et al. (117), we consider that we broadened the scope of the review, moving from the home telemonitoring and structured telephone support to the more comprehensive concept of early warning system. This explains the five extra papers comprised in our review, including the one by Pandor et al. (117). Three of the additional studies focused on some type of B-type natriuretic peptide measurement (113, 115, 116), while the remaining one focused on remote cardiac implantable electric devices management in patients implanted with implantable cardioverter defibrillators (111).

We opted for a broader scope because we believe that all studies abiding to the suggested definition of early warning systems have common points on their interventions and can be evaluated under the same label. For this reason we wanted to test whether they can be modelled using similar approaches. Despite finding some differences between the models used in the various types of interventions, we also found that our line of reasoning was correct, as there are clear similarities amongst the models used in all seven studies.

Finally we would like to praise highly the study by Goehler et al. (129), which systematically reviewed and assessed the modelling approaches with the criteria of the German Competence Network on Heart Failure in mind. It partially overlaps with the present study: two out of the seven papers included in the present review were also included in their work, while four of the five unreported papers were yet to be published at the time of their study; still, their exclusion of the study by Klersy et al. (112) remains unjustified. Notwithstanding, it is fair to recognize that the paper by Goehler et al. (129) was not superseded by the current review, as its focus was not restricted to early warning systems and to the outpatient setting.

Lessons learned

It was beyond the objectives of our review to provide a final recommendation as to which model approach should be preferred. Modelling is an *ad hoc* exercise which greatly depends on research objectives, target population, health technologies evaluated, and availability of data. However, it has been advocated that standardized disease-specific models should be developed as a way of decreasing structural uncertainty amongst models used for the economic evaluation of that same disease (130). In the case of heart failure taking into account the time parameter is paramount in order to monitor the eventual progression of the disease. Accordingly, Markov and discrete event simulation models seem more appropriate for modelling in the field of heart failure management. Notwithstanding the technical issues that may arise in the development of these models, these are usually well understood by clinicians and they benefit of feeding on traditional epidemiological survival data such as annual rates, Kaplan-Maier curves, time-to-event distributions (129). We should be sceptical

towards the use of decision trees in HF management modelling. In a field that deals with a chronic disease, one that may develop over several decades, these trees would turn out to be over-simplistic, lest they become prohibitively complicated.

In the case of Markov modelling, two different types of health states were used: disease progression state (severity of disease, i.e. NYHA class) or health care provision state (hospitalisation status). Albeit dependant on patient-reported outcomes, NYHA class is widely accepted by cardiologists as an accurate measurement of disease severity (12). Thus, given that HF is a multi-dimensional disease lacking a unique biomarker for measuring disease severity, it seems more appropriate to model early warning systems for the management of heart failure – which could be best evaluated by focussing on the deterioration patient health status – using NYHA class as Markov states accounting for disease severity rather than using the number of previous hospitalisations as a proxy. Nonetheless, it should be taken into account that this may not always be possible due to the outcomes measured in the randomized controlled trials (RCTs) that originate the cost-effectiveness analysis or that are used as data sources for the study. In the particular case of the object of our review, it appears to have been the case for Morimoto et al. (116) and Moertl et al. (115) choosing hospitalisation status as health states and for Miller et al. (114) opting for using NYHA classes in its structure.

Still regarding the constraints imposed by RCTs on economic evaluations, one of the core issues in modelling has to do with the extrapolation of the treatment effect from the RCT period to the time horizon of our evaluation. Extrapolation mainly depends on the outcomes measured in the RCT and which of these outcomes will be used to estimate the difference in efficiency between the treatment and the comparator. However, if patient-level data is available, extrapolation by means of survival analysis – using both parametric and non-parametric models – is preferred. Alternatively, if only counts for hospitalisations and/or NYHA class are available, transition probabilities should be calculated.

Finally, some considerations about modelling hypothetical scenarios when no hard data is available or when the right comparator is absent. In this situation – assuming that decision trees are not viable alternatives for modelling early warning systems in the management of HF – hazard ratios for hospitalisation and/or mortality should be employed. This is both valid for Markov models and discrete event simulation models. Further, in discrete event simulation modelling it is possible to explore the impact of changes in specific patient characteristics on the outcomes of the model. In the particular case of early warning systems, this feature of discrete event simulation modelling is expected to be of vast interest. Since the observed variation in the collected data for individual patients is what will be responsible for triggering alarms

and follow-up actions, explaining this variation in terms of its cost-effectiveness will be crucial for determining the threshold for raising eventual alarms.

Limitations

The present study acknowledges two major limitations. The first derives from the fact that only models based on published full-text articles were assessed. Additional information (in particular, technical modelling reports) via direct communication with the authors was not requested. Since full-text papers seldom display all available information relating to their models, it is possible that one or more items that might be of interest remain unreported. This may directly lead to bias on the scoring of the Philips checklist. Handels et al. (131) reported this problem while reviewing the economic evaluations of interventions for the early diagnosis of Alzheimer's disease and related disorders. Those authors suggest to overcome this lack of information by including extensive appendixes, and they point to Getsios et al. (132) and, in a different field, to Van Gestel et al. (133) as valid alternatives for publications of decision-analytical models.

The aforementioned second limitation relates to the definition of early warning systems. This concept covers a plethora of health technologies and their associated interventions, with the concomitant inter-variability within technologies. To overcome this limitation a sensitive search strategy was adopted in order to enhance the retrieval probability of relevant studies.

Recommendations

Keeping an eye on the main flaws so far identified in the current models, two main points should be stressed.

Firstly, we posit that further research on early warning systems, a field with a huge potential, waits to be done. A comprehensive and well acknowledged definition for early warning systems is highly desirable, as they are certainly about to become a conspicuous fixture in health research, reflecting the strong need to move from curative care to preventative care.

And secondly, we point to the desirability for authors making technical information regarding the models more explicit in their publications. As Rennie and Luft (134) point out, in contradistinction to clinical studies, which tend to focus on the consequences of an intervention, economic evaluations demand more reporting space for additional items like resource use, costs, preference related information, and cost-effectiveness results. Their lacking – often a consequence of the limited word count available for paper publication – is a major drawback for editors, reviewers, and researchers wishing to scrutinize the studies' features. The Consolidated Health

Economic Evaluation Reporting Standards (CHEERS) statement, published by Husereau et al. (135), provides recommendations, in the form of a checklist, to optimize reporting of health economic evaluations. The use of that checklist for the sake of transparency and for the simplification of interpretations and comparisons should be strongly recommended.

Conclusions

A systematic review of decision-analytical models used in the economic evaluation of early warning systems for the management of chronic heart failure patients was conducted, retrieving seven modelling studies. Some variability was found with regards to their general characteristics and methodological quality. Particularly, some quality features were not properly addressed in the reviewed studies or insufficient information was available in order to assess them appropriately.

Further research waits to be done, starting preferentially with the development of a comprehensive definition of early warning systems and with the extension of work to other chronic diseases, chronic obstructive pulmonary disease (COPD) and diabetes being the foremost candidates. Furthermore, future modelling exercises should describe in detail any competing theories regarding model structure and disease progression. Models should also be more precise with regards to the identification of key parameters and the use of expert opinion, and should ensure that the four types of uncertainty (methodological, structural, heterogeneity, and parameter) are properly assessed. Finally, the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist should be adopted for reporting the methods and findings of future economic evaluation studies for the sake of clear-sighted interpretation and comparison of the studies under consideration.

Expert commentary

Early warning systems may be defined as health technologies that consist mainly of three differentiating elements: (i) monitoring and collection of clinical data (e.g., vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a particular combination of signs and symptoms – that trigger an alarm and follow-up actions.

The current body of evidence concerning health and economic outcomes of early warning systems is inconclusive. The complexity of these devices – as a result of their widespread applications and potentialities – makes them a challenging research topic that has yet to receive enough attention by the scientific community. New research methods for collecting, analysing and interpreting data about early warning systems are required.

Modelling techniques have become quintessential in the economic evaluation of new health technologies. Although it can be argued that the results of these studies are not the single motive for making decisions on the adoption of innovative technologies, they are certainly a criterion to which decision-makers pay substantial attention. For this reason, it is key to develop methodologies for health economic modelling in the field early warning systems that take into account the specific characteristics of these systems, particularly the monitoring and collection of clinical data. While monitoring systems become more complex and allow for more frequent and repeated measurements, a challenge is posed for health economic modelling. Data generated by these measurements implies that modelling techniques take into account individual patient characteristics, as well as their changes over time. Discrete event simulation is a type of modelling used in health technology assessment that allows for using patient characteristics as determinants for health and economic outcomes. Due to the technical challenges in their development and their huge demand for data, these models are not common practice in health technology assessment research. However, the findings of this research substantiate the use of discrete event simulation models in cost-effectiveness studies in heart failure in general – because of the multi-dimensional characteristics of the disease – and in early warning systems in particular, as it seems to be the best methodological approach for assessing the impact of the variation of individual patient characteristics on patient outcomes both in terms of costs and effects.

Five-year view

Health policy is increasingly driven by evidence published in economic evaluations and other evidence-based studies, thereby resulting in higher requirements for the methodologies used in these studies. Economic evaluations that do not include any kind of decision-analytic modelling are now seldom published in peer-reviewed journals.

Early warning systems and other interventions relying on technology to facilitate preventive medicine are now in vogue in the most developed health care systems in the world. The future of these technologies will be dictated by the upcoming innovation in this field, the empowerment of patients – by involving them in their own health care –, and the acceptability of innovative monitoring techniques by health care professionals and patients. However, as it seems unquestionable that the rise in chronic disease prevalence will increase the demand for innovative solutions, early warning systems are expected to play a great role in shaping the future of health care. Further, they are expected to expand both in number and in their technical features.

In the next 5 years, readers can expect to see more studies focusing on the cost-effectiveness of early warning systems. As the scientific community must strive for generating transparent and reliable evidence that leads decision-makers into making right choices regarding the allocation of resources in health care, it is crucial to develop more robust methods for modelling in this field. For this purpose, modellers must combine and adapt lessons learned from previous modelling endeavours and create a culture of scientific openness and cooperation. Furthermore, understanding modelling approaches used in early warning systems for the management of heart failure can make way to the use of these models in other chronic diseases like diabetes and chronic obstructive pulmonary disease.

Key issues

- The most frequently used methodological approaches in decision-analytic modelling do not seem adequate for the assessment of the cost-effectiveness of new emerging technologies like early warning systems
- Describing the different approaches used in published decision-analytic models for the management of heart failure management using early warning systems is expected to provide concise and useful insight to inform the future development of models in this field.
- Variability was found with regards to the general characteristics and methodological quality of reviewed decision-analytic models; some quality features were not properly addressed in the reviewed studies or insufficient information was available in order to assess them appropriately.
- Future modelling exercises should describe in detail any competing theories regarding model structure and disease progression. Models should also be more precise with regards to the identification of key parameters and the use of expert opinion, and should ensure that the four types of uncertainty (methodological, structural, heterogeneity, and parameter) are properly assessed.
- The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist should be adopted for reporting the methods and findings of future economic evaluation studies for the sake of clear-sighted interpretation and comparison of the studies under consideration.
- In order to fully grasp the potential value of early warning systems for patients and patient outcomes, modelling studies should take patient characteristics into account. In this instance, discrete event simulation models should be preferred.

Appendices

Appendix 2.1 – EMBASE search strategy

1. 'heart failure'/exp
2. ((cardi* OR heart*) NEAR/3 (fail* OR insuffic* OR incompeten*)):ab,ti
3. 1 OR 2
4. 'economic evaluation'/exp
5. 'quality adjusted life year'/exp
6. (cost* NEAR/3 (effective* OR benefit* OR utilit* OR minimi*)):ab,ti
7. ((economic* OR pharmacoeconomic*) NEXT/1 evaluation*):ab,ti
8. (value* NEAR/2 (money OR monetary)):ab,ti
9. ((decision* OR 'individual patient' OR 'individual patients' OR cohort) NEXT/3 model*):ab,ti
10. markov*:ab,ti
11. 'decision tree':ab,ti
12. 'decision trees':ab,ti
13. (('discrete event' OR 'discrete events' OR patient) NEXT/1 simulat*):ab,ti
14. ('quality adjusted' NEAR/3 ('life year' OR 'life years')):ab,ti
15. qaly*:ab,ti)
16. OR/4-15
17. 3 AND 16
18. 'early diagnosis'/exp
19. 'prevention'/exp
20. 'prediction'/exp
21. 'preventive health service'/exp
22. ‘case management’/exp
23. ‘clinical protocol’/exp
24. ‘patient care planning’/exp
25. 'disease course'/exp
26. 'clinical pathway'/exp
27. 'home care'/exp
28. 'outpatient'/de
29. 'outpatient care'/de
30. 'ambulatory care'/exp

31. ‘patient monitoring’/exp
32. 'community care'/exp
33. ‘hospital admission’/exp
34. ‘hospital readmission’/exp
35. 'telehealth'/exp
36. 'telemetry'/exp
37. 'telephone'/exp
38. ‘teleconsultation’/exp
39. 'health program'/exp
40. ((earl* NEAR/3 (warn* OR diagnos* OR detect* OR intervent* OR alarm*)):ab,ti
41. ((warn* OR detect* OR alarm* OR monitor*) NEAR/3 (system*)):ab,ti
42. ((implantable or wearable) NEAR/3 (system* OR sensor*)):ab,ti
43. prevent*:ab,ti
44. predict*:ab,ti
45. ‘case management’:ab,ti
46. (clinical NEAR/3 protocol*):ab,ti
47. (‘patient care' NEAR/3 planning*):ab,ti
48. ((disease* OR clinical*) NEAR/3 (course* OR pathway* OR management*)):ab,ti
49. progress*:ab,ti
50. deteriorat*:ab,ti
51. prognos*:ab,ti
52. innovati*:ab,ti
53. ((home OR communit* OR domicilliar* OR ambulator* OR remote* OR patient* OR outpatient* OR tele) NEAR/3 (care OR monitor* OR management OR consultat*)):ab,ti
54. ((hospital* OR patient* OR voluntar*) NEAR/3 (admission* OR readmission*)):ab,ti
55. Telehealth:ab,ti
56. Ehealth:ab,ti
57. 'mobile health':ab,ti
58. mhealth*:ab,ti
59. telephone*:ab,ti
60. telemetr*:ab,ti
61. teleconsult*:ab,ti
62. telemonitor*:ab,ti
63. ((health OR healthcare) NEAR/3 program*)):ab,ti

64. OR/18-63
65. 17 AND 64
66. ([animals]/lim NOT [humans]/lim) NOT ([conference abstract]/lim OR [letter]/lim OR [note]/lim OR [conference paper]/lim OR [editorial]/lim OR [conference review]/lim) AND [english]/lim
67. 65 NOT 66

Appendix 2.2 – Excluded studies, with reasons.

Author, year	Reason for exclusion
1. Adlbrecht, C; et al., 2011	Not an early warning system, no decision-analytical model is present
2. Bentkover, JD; et al., 2003	Not an early warning system, not a cost-effectiveness study
3. Bocchi, EA; et al., 2013	Conference abstract
4. Burri, H; et al., 2013	Included; it should be formally excluded because population under study concerned patients with cardiac implantable electric devices, and was not restricted to the ones diagnosed with heart failure
5. Cano Martin, JA; et al., 2014	Not an early warning system
6. Chan, DC; et al., 2008	Not an early warning system
7. Chen, Q; et al., 2000	Not heart failure, not an early warning system, no decision-analytical model is present
8. Cui, Y; et al., 2013	No decision-analytical model is present
9. Gohler, A; et al., 2008	Not an early warning system, not outpatient setting
10. Graves, N; et al., 2009	Not an early warning system
11. Hailey, D; Yu, P., 2013	No decision-analytical model is present
12. Heidenreich, PA; et al., 2004	Not an early warning system, not outpatient setting
13. Henderson, C; et al., 2013	No decision-analytical model is present
14. Henderson, C; et al., 2013	No decision-analytical model is present (same study as 13.)
15. Inglis, SC; et al., 2006	Not a cost-effectiveness study
16. Milburn, AB; et al., 2014	Not a cost-effectiveness study
17. Noel, HC; et al. 2004	Not an early warning system, no decision-analytical model

	is present
18. Roth, A; et al., 2006	Not a cost-effectiveness study
19. Thokala, P; et al., 2013	Decision-analytical model is the same as the one presented in the included study: Pandor, A; et al., 2013
20. Thokala, P; et al., 2013	Conference abstract (refers to study 19.)
21. Zanaboni, P; et al., 2013	No decision-analytical model is present

Appendix 2.3 – Narrative description of reviewed studies

Burri et al. 2013. Burri et al. (111) performed a cost-consequence analysis (a variation of the cost-effectiveness approach that provides costs and outcomes in disaggregated form) assessing the long-term costs and consequences of using remote cardiac implantable electric devices management in patients implanted with implantable cardioverter defibrillators and cardiac resynchronization therapy defibrillators for either primary or secondary prevention. The authors used a Markov model with a time horizon of 10 years and a cycle length of 1 year. Four main health states were included: Well, Poststroke, Post-ADHF, and Dead as an absorbing state. In every Markov cycle and health state (except 'Dead'), the probabilities of experiencing the clinical and technical events were applied. The role of modelling was to inform decisions in the absence of hard data. Twelve consequences were examined in the model: scheduled and unscheduled FU visits, battery replacements, lead malfunctions, atrial fibrillation/flutter (AF), lead-related inappropriate shocks, non-lead-related inappropriate shocks, stroke, hospital admission for acute decompensated heart failure (ADHF), sustained ventricular arrhythmia (SVA), appropriate shocks triggered by SVA, and death. Deterministic sensitivity analysis of model parameters was performed and presented with a tornado diagram. The main conclusion of this study was that the model establishes HM as an economically viable technology when applied within the UK NHS system.

Klersy et al. 2011. Klersy et al. (112) assessed the cost-effectiveness and the cost-utility of remote patient monitoring when compared with the usual care approach based upon differences in the number of hospitalizations, estimated from a meta-analysis of randomized clinical trials. The authors used a decision tree and budget impact model with a time horizon of 1 year. The decision tree in both usual care and RPM patients considered two options: (i) the patient is hospitalized for HF (event); or (ii) the patient is not hospitalized for HF (event free) during follow-up. The budget impact model simulated the economic impact of a change in the approach of the care pathway of a theoretical population of HF patients followed-up for 1 year with an RPM

implementation rate ranging from 0 to 50%. The role of modelling consisted in synthesizing head-to-head comparisons where relevant trials do not exist. The authors came to the conclusion that the novel cost-effectiveness data coupled with the demonstrated clinical efficacy of remote patient monitoring should encourage its acceptance amongst clinicians and its consideration by third-party payers.

Laramée et al. 2013. Laramée et al. (113) calculated the cost-effectiveness of three strategies for optimising medical therapy in patients with chronic heart failure: management guided by serial measurement of circulating natriuretic peptide concentration by a specialist, management guided by clinical assessment by a specialist, and usual care in the community. For the purpose of informing decisions in the absence of hard data, the authors built a decision tree. The decision tree for intervention was sub-divided in True HF and No HF, which respectively sub-divided in BNP above threshold, echo positive and BNP below threshold, HF missed, and BNP above threshold, echo negative and BNP below threshold, no echo. The authors concluded that the optimisation of medical therapy in CHF guided by serial natriuretic peptide measurements by a specialist is cost-effective – at a threshold of £20 000 per QALY and only for specific subgroups of patients – when compared with both medical therapy guided by specialist's clinical assessment and usual care in the community.

Miller et al. 2009. Miller et al. (114) performed a cost-utility analysis using a Markov model to extrapolate from the results of an 18-month clinical trial to estimate the long-term impact of systolic heart failure disease management when compared with usual care. The model had a lifetime time horizon and a 6-month cycle length. Patients represented in the model can transition among 3 disease states based on NYHA class (class I, class II, and a grouping of classes III and IV), with a possible transition to death from each of these states. The results of the study show that discounted lifetime program and medical costs were $4850 higher in the disease management group and that the intervention had a long-term discounted cost-effectiveness of $43,650/QALY. Deterministic sensitivity analysis was performed as a way of assessing uncertainty.

Moertl et al. 2013. Moertl et al. (115) assessed the cost-utility of NT-proBNP-guided intensive patient management on top of multidisciplinary care, when compared with multidisciplinary care alone or usual care. The authors used a Markov model with a 20-year time horizon and a 1-month cycle length to extrapolate the results of a 18-month clinical trial. The model simulated disease progression from HF using the number of previous HF hospitalizations as a proxy for disease progression. Health states within the model were representative of the number of previous HF hospitalizations with the assumption of greater risks of subsequent hospitalizations and mortality the greater the number of previous hospitalizations. The study concluded that NT-proBNP-guided intensive heart failure patient management in addition to multidisciplinary care was the most cost-effective strategy, as it was

dominant over both multidisciplinary care alone or usual care. Probabilistic sensitivity analysis was performed to show that the probabilities for NT-proBNP-guided intensive heart failure patient management in addition to multidisciplinary care being the most cost-effective strategy were 92% at a threshold value of Austrian €40,000 and 93% at a threshold value of Canadian $50,000.

Morimoto et al. 2004. Morimoto et al. (116) used a Markov model for informing decisions in the absence of hard data and assessing the cost-utility of chronic heart failure management with B-type natriuretic peptide measurement when compared to HF management without B-type natriuretic peptide measurement. The model had a 10-year time horizon a 3-month cycle length. Patient enter the model in the state of "no additional hospitalization for CHF" and every 3-month cycle can remain in the same health state or move along 4 other health states: "1st additional hospitalization for CHF", "2nd additional hospitalization for CHF", "3rd additional hospitalization for CHF", or "4th or more additional hospitalizations for CHF". From every of the five health states patients can move to health state "dead". Baseline analysis of this study revealed dominance of the group with B-type natriuretic peptide measurement. Deterministic sensitivity analysis was the only type of uncertainty assessment performed in this study.

Pandor et al. 2013 Pandor et al. (117) determined the cost-effectiveness of home telemonitoring or structured telephone support strategies when compared with usual care for adult patients who have been recently discharged (within 28 days) from acute care after a recent exacerbation of HF. The authors developed a comprehensive Markov model synthesizing head-to-head comparisons where relevant trials do not exist, and informing decisions in the absence of hard data. The model had with a 30-year tome horizon and a cycle length of 1 month. Two different states were considered: (a) alive at home and (b) dead. In each period the patients who were alive were under the risk of an average number of monthly rehospitalisations; each patient then accrued lifetime QALYs and health care costs according to their hospitalisation and treatment status. The model was subject to scenario analysis for costs, probabilistic sensitivity analysis for all model parameters (10,000 simulations), and expected value of perfect information (EVPI) analysis. The study concluded that home telemonitoring during office hours was the most cost-effective strategy with an estimated incremental cost-effectiveness ratio (ICER) of £11,873 per quality-adjusted life-year (QALY) compared with usual care.

Chapter 3

Impact of hospitalisation on health-related quality of life in patients with chronic heart failure

Fernando Albuquerque de Almeida, Maiwenn J. Al, Ron Koymans, Jarno Riistama, Steffen Pauws and Johan L. Severens

Health and Quality of Life Outcomes (2020) 18:262.

Abstract

Background: Empirical identification of the direct impact of hospitalisation in the change in utility could provide an interpretation for some of the unexplained variance in quality of life responses in clinical practice and clinical trials and provide assistance to researchers in assessing the impact of a hospitalisation in the context of economic evaluations. This study had the goal of determining the impact of nonfatal hospitalisations on the quality of life of a cohort of patients previously diagnosed with heart failure by using their quality of life measurements before and after hospitalisation.

Methods: The impact of hospitalisation on health-related quality of life was estimated by calculating the difference in utility measured using the EQ-5D-3L in patients that were hospitalised and had records of utility before and after hospitalisation. The variation in differences between the utilities pre and post hospitalisation was explained through two multiple linear regression models using (1) the individual patient characteristics and (2) the hospitalisation characteristics as explanatory variables.

Results: The mean difference between health-related quality of life measurement pre and post hospitalisation was found to be 0.020 [95% CI: -0.020, 0.059] when measured with the EQ-5D index, while there was a mean decrease of -0.012 [95% CI: -0.043, 0.020] in the utility measured with the visual analogue scale. Differences in utility variation according to the primary cause for hospitalisation were found. Regression models showed a statistically significant impact of body mass index and serum creatinine in the index utility differences and of serum creatinine for utilities measured with the visual analogue scale.

Conclusion: Knowing the impact of hospitalisation on health-related quality of life is particularly relevant for informing cost-effectiveness studies designed to assess health technologies aimed at reducing hospital admissions. Through using patient-level data it was possible to estimate the variation in utilities before and after the average hospitalisation and for hospitalisations due to the most common causes for hospital admission. These estimates for (dis)utility could be used in the calculations of effectiveness on economic evaluations, especially when discrete event simulations are the employed modelling technique.

Background

Heart failure (HF) is a condition characterized by typical symptoms (e.g., breathlessness, ankle swelling and fatigue) and signs (e.g., elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress (12). HF is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. From a medical perspective, the goals of managing patients with HF consist in improving their clinical status, functional capacity, and quality of life, preventing hospital admission, and reducing mortality (10, 12). Understanding the relationship between all these goals of HF therapy is of vital importance for informing the development of clinical practice guidelines and for approving or recommending new therapeutic interventions for HF.

Previously published studies indicate that quality of life or health-related quality of life (QoL or HRQoL, respectively; henceforth used interchangeably) in patients with HF is greatly impaired when compared to the general population (23, 24, 27). The New York Heart Association (NYHA) functional classification – a system for classifying patients according to the severity of their symptoms – has been shown to be a strong independent predictor of QoL for patients with HF (27). However, since NYHA functional class is only assessed during clinical visits and provides a relatively simplistic way of classifying the extent of heart failure based on patients' limitations during physical activity, the underlying determinants of reduced quality of life in patients with HF remain hardly distinguishable (136), particularly HF-related events that are expected to have an impact on patient utility (e.g., hospitalisation) (137).

Economic evaluations published in the literature have used an estimated disutility for hospitalisation equivalent to the decrease in utility between a particular NYHA class and the one immediately worse (138). In view of the high incidence of (re)hospitalisation in patients suffering from HF, in absence of a robust method for calculating the (dis)utility resulting from a hospitalisation, it becomes essential to explore the relationship between hospitalisation and quality of life. From a theoretical viewpoint it may be assumed that there is a relationship between hospitalisation and utility, although there is insufficient or unclear reporting of evidence about the impact on utility caused by the hospitalisation of HF patients, both in the magnitude of the effect and the duration of this same effect (117).

Current practice in economic evaluations aimed at estimating quality-adjusted life years (QALYs) consists in measuring utility at specific points in time and linearly interpolate these values so that they reflect a larger time period for the subjects under analysis. In others words, using the QALY model as the measure of effectiveness in

economic evaluations implies missing temporary changes in utility, particularly when these changes are due to disease-related events. For instance, when having access to two consecutive utility measurements with the same value – one before and other after a particular event –, using the QALY model leads to an implicit assumption that the utility of that same patient was constant throughout both time points and that the event that took place had no influence in the QoL of that patient, even though this assumption is unlikely to hold in an event such as a hospitalisation (139, 140). In this sense, empirical identification of the direct impact of hospitalisation on the change in utility could provide an interpretation for some of the unexplained variance in QoL responses in clinical practice and clinical trials, as well as it may provide assistance to researchers in designing trials aimed at assessing patient-reported outcomes.

This study had the goal of determining the impact of nonfatal hospitalisations on the QoL of a cohort of patients previously diagnosed with heart failure by using their QoL measurements before and after hospitalisation.

Methods

Data

We used the data from the Trans-European Network-Home-Care Management System (TEN-HMS) trial for our study. This trial investigated the impact of using home telemonitoring, nurse telephone support (NTS), or usual care (UC) in hospital admissions, hospital days, and rates of mortality. Details of the inclusion and exclusion criteria, follow-up, and results of the study have been reported elsewhere (126). In brief, patients who were ready for discharge or who were recently discharged after an hospital admission due to heart failure were evaluated for inclusion conditional on the permission by their primary care physician. Inclusion criteria for patients consisted of a hospital admission due to or complicated by worsening heart failure lasting more than 48 hours within the last six weeks, persisting symptoms of heart failure, LV (left-ventricular) ejection fraction <40%, LV end-diastolic dimension >30 mm/m (height), and being medicated with furosemide at a dose ≥40 mg/day or equivalent (e.g., ≥1 mg of bumetanide or ≥10 mg of torasemide). In addition, patients should have at least one of the following indicators of further increase in risk: (1) unplanned cardiovascular admission lasting more than 48 hours within the previous 2 years; (2) LV ejection fraction <25%; or (3) treatment with furosemide at a dose of ≥100 mg/day or equivalent. Patients younger than 18 years of age who were considered incapable of complying with home telemonitoring or who were awaiting revascularisation, cardiac resynchronisation, or heart transplantation were excluded.

Theoretical framework

The health-related quality of life in heart failure depends on the specific characteristics of a given patient, such as the disease status, gender, comorbidities, age, among others (24). Events that may alter any of the aforementioned characteristics are expected to have an indirect impact on HRQoL. Because it results from a temporary deterioration of the health status of the patient or a permanent change in health status deriving from the progression of the disease, being hospitalised is expected to have an effect in HRQoL.

In practice, utility measurements of HF patients are taken periodically, during clinical visits to the physician, and they are not always performed when particular events related to disease progression take place (e.g., pulmonary embolism, tachyarrhythmia, hospitalisation). Hence, while the global trend in HRQoL can be summarised, the specific impact of the event may be concealed, leaving many associations that can be hypothesised. For instance, in the period before the event, QoL may be decreasing as a result of a decline in the health status of the patient – which may in part explain hospitalisation –, but after the event QoL may improve again. As a result, the

difference between the last utility measurement pre-event and the first measurement post-event could be zero or even show an increase in QoL. Alternatively, QoL may be stable when a very sudden decline in health triggers the event. After this event, QoL may improve again but it may not get back to the level it was before the decline happened. In this particular situation a decrease between the last pre-event utility measurement and the first post-event utility measurement would be recorded.

Using the data from the TEN-HMS trial we will try to answer our research question by analysing differences between QoL before and after hospitalisation. In this way we will be able to infer on the hypothesis that there is a difference in utilities resulting from the hospitalisation event. This approach entails that for every considered hospitalisation there is a period of time pre and post event that may vary for every observation and that may result in a different magnitude of the utility change between both measurements. Moreover, the particular characteristics of the hospitalisation – length of stay and whether the patient was admitted to the intensive care unit – are also likely to have an influence on the variation of HRQoL. Figure 3.1 provides a schematic representation of the framework that will be used for testing the hypothesis that hospitalisation impacts HRQoL and the determinants that may play a role in the measured variation.

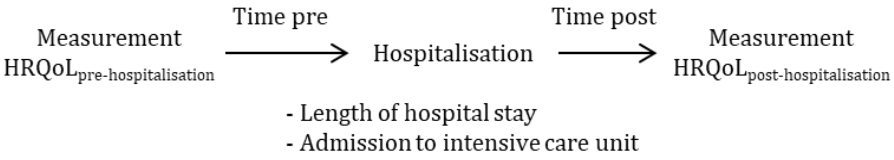


Figure 3.1 – Schematic representation of HRQoL and its determinants

Measurement of the health-related quality of life

In the TEN-HMS trial health-related quality of life was measured using the three-level EQ-5D questionnaire (henceforth EQ-5D-3L), which consists of a descriptive system and a visual analogue scale (VAS) (141, 142). The validity and reliability of the EQ-5D tool as an outcome measure within the cardiovascular area have been previously asserted (143). More specifically, it has shown satisfying psychometric properties in cardiac rehabilitation (144).

For this study we calculated utilities by applying the utility weights previously identified for the Netherlands to the answers given to the EQ-5D-3L descriptive system questionnaire (utility values found this way will be referred to as index utilities, as opposed to VAS utilities) (145).

The EQ-5D-3L in the TEN-HMS trial was administered at baseline and it was repeated at 4, 8, 12, 16, 20, and 24 months, during scheduled clinical visits.

Hospitalisation

There were sixty different causes for admission identified in the dataset. From these, hospitalisations could be classified into three major groups: (1) due to heart failure, (2) other cardiovascular, or (3) noncardiovascular. Information regarding the number of days spent in hospital, whether the patient was admitted to the ICU, and if the patient died during hospitalisation were also available from the data.

Statistical analyses

HRQoL pre and post hospitalisation (base case)

In order to assess the impact of hospitalisation on HRQoL, we took the available EQ-5D-3L measurements immediately before and after hospitalisation. We then calculated the difference in utility measured for each individual patient – both using index and VAS utilities –, followed by the average utility difference for all patients who were hospitalised and had records for both measurements.

Sensitivity analyses

Four sensitivity analyses were performed. First, we excluded patients who experienced more than one hospitalisation between the EQ-5D-3L assessments of interest. Second, we only considered the hospitalisations for which the reason for admission was either heart failure or other cardiovascular event. Third, we stratified patients into consecutive groups for those who completed the EQ-5D-3L within X days of the non-fatal hospitalisation (for X = 20, 40, 60, 80, 100), in order to determine whether the time interval between the event and the subsequent HRQoL assessment had any effect on the magnitude of the utility change. And finally, we performed an analysis in which patients that died after hospitalisation and before completing the following HRQoL assessment were assigned a value of 0 for their utility measurement.

Utility variation by primary admission cause

In order to infer on the impact of the most frequent events that can lead to hospitalisation on utilities of HF patients, we used the methods from the base case analysis individually for each of the ten most common reasons for primary admission described in the dataset.

Impact of the characteristics of the patient and of the hospitalisation in the variation in HRQoL

We aimed at explaining the variation in differences between the utilities pre and post hospitalisation through two multiple linear regression models. The first used individual patient characteristics (measured at the same moment as utilities) as explanatory variables: body mass index, systolic and diastolic blood pressures,

haemoglobin, serum sodium, and creatinine; the second used hospitalisation characteristics as the explanatory variables: length of hospital stay, number of days between the measurement before hospitalisation, number of days elapsed between the considered hospitalisation and the subsequent utility measurement, and a binary variable for the admission to the intensive care unit (ICU).

All statistical analyses were conducted using the programming language R (146).

Results

Baseline data

The demographic and clinical characteristics at baseline for the total population (n = 426) and the sub-population which has been hospitalised at least once can be found in Table 3.1. From the total population included in the study, 270 individuals (63.4%) experienced at least one hospitalisation (total number of hospitalisations = 583); the data from these patients were used in the analyses.

The average age of included patients was 67.1 years old and there is a 4:1 ratio of men over women in this population. The great majority of patients have comorbidities, especially previous myocardial infarction and hypertension. Previous myocardial infarction is the main primary cause for HF in 63.3% of the cases, followed by idiopathic dilated cardiomyopathy (20.0%). Higher utilities at baseline (with higher standard deviation) were recorded for index utilities when compared to VAS utilities (0.669 ± 0.246 vs. 0.537 ± 0.189, respectively).

Table 3.1 – Baseline Characteristics

Variable	**Total population**	**Hospitalised population**
Number	426	270
Hospitalised at least once (%)	270 (63.4)	270 (100)
Mean age, years (SD)	67.1 (13.1)	67.1 (13.2)
% patients age ≥70 years	48.1	48.5
% Women	22.5	19.6
Lives alone (%)	113 (26.5%)	69 (25.6%)
Lives with partner or friend (%)	313 (73.5%)	201 (74.4%)
Primary cause of heart failure (%)		
Coronary disease	254 (59.6%)	171 (63.3%)
Hypertension	27 (6.3%)	15 (5.6%)
Idiopathic dilated cardiomyopathy	95 (22.3%)	54 (20.0%)
Alcohol-related	11 (2.6%)	4 (1.5%)
Valve-related	28 (6.6%)	20 (7.4%)
Other	10 (2.3%)	6 (2.2%)
***Comorbidities* (%)**		
Previous myocardial infarction	241 (57%)	163 (60%)

Valve disease /mitral regurgitation	156 (37%) /138 (32%)	101 (37%) /82 (30%)
Chronic or paroxysmal atrial fibrillation	192 (45%)	127 (47%)
Hypertension	200 (47%)	123 (46%)
Stroke, any	39 (9%)	29 (11%)
Chronic lung disease	103 (24%)	69 (26%)
Diabetes, any	149 (35%)	94 (35%)
Investigations (SD)		
Weight (kg)	76.7 (16.7)	77.1 (16.6)
Body mass index (kg/cm^2)	26.2 (4.7)	26.3 (4.8)
Systolic blood pressure (mm Hg)	114.2 (19.3)	113.1 (19.7)
Diastolic blood pressure (mm Hg)	69.3 (11.3)	69.1 (11.3)
Haemoglobin (g/dl)	13.0 (2.1)	12.9 (2.0)
Serum sodium (mmol/l)	137.5 (5.0)	137.3 (5.1)
Serum creatinine (μmol/l)	138.7 (54.0)	143.7 (58.7)
Mean LVEF (%)	26.0 (7.5)	26.1 (7.7)
% with LVEF <25%	50.2	47.4
NT proBNP (pg/ml), median [IQR]	365.5 [152.3 to 796.5]	393.0 [177.5 to 871.0]
Utility – Index (SD)	0.687 (0.242)	0.669 (0.246)
Utility – VAS (SD)	0.538 (0.192)	0.537 (0.189)

Abbreviations: IRQ, interquartile range; LVEF, left ventricular ejection fraction; SD, standard deviation.

Statistical analyses

HRQoL pre and post hospitalisation (base case)

The mean difference between the HRQoL measurement pre and post hospitalisation was found to be 0.020 [95% CI: -0.020, 0.059] for index utilities, and -0.012 [95% CI: -0.043, 0.020] for VAS utilities. There were no striking differences between the shape of the density curves of the utility variation when measured with either the EQ-5D-3L index or the VAS (cf. Figure 3.2).

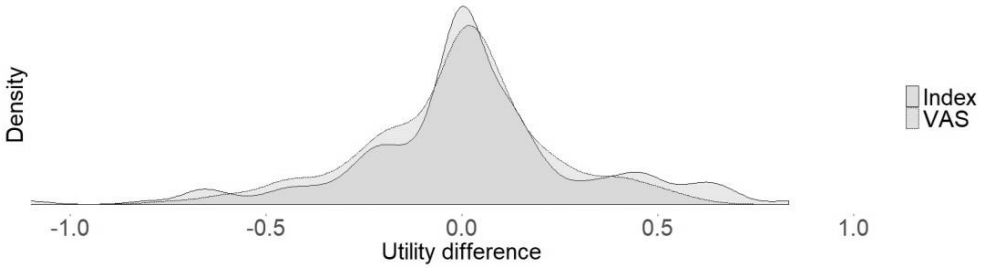


Figure 3.2 – Distribution of the difference between measurements pre and post hospitalisation for index and VAS utilities

Sensitivity analyses

The first sensitivity analysis consisted of excluding patients who had more than one hospitalisation but did not die before the following HRQoL assessment. The analysis provided consistent results when compared to the primary analysis: utility variation of 0.000 [95% CI: -0.081, 0.081] with index utilities and -0.019 [95% CI: -0.084, 0.046] with VAS. Secondly, restricting the analysis to hospitalisations that were due to cardiovascular conditions alone also does not change results of QoL variation substantially, with a calculated increase in utility of 0.023 [95% CI: -0.016, 0.062] for index utilities and a decrease of -0.009 [95% CI: -0.041, 0.023] for VAS. Thirdly, stratifying patients according to the number of days elapsed between hospitalisation and the subsequent utility measurement, despite the large variance, shows that differences in utility measured with the VAS are noticeably smaller in absolute terms when compared to index utility differences (see Figure 3.3). And finally, when assigning 0 to the utility score of patients who died after the hospitalisation, there was a significant decrease in the utilities pre and post hospitalisation of -0.172 [95% CI: -0.222, 0.122] with the EQ-5D index and -0.133 [95% CI: -0.171, -0.096] for VAS.

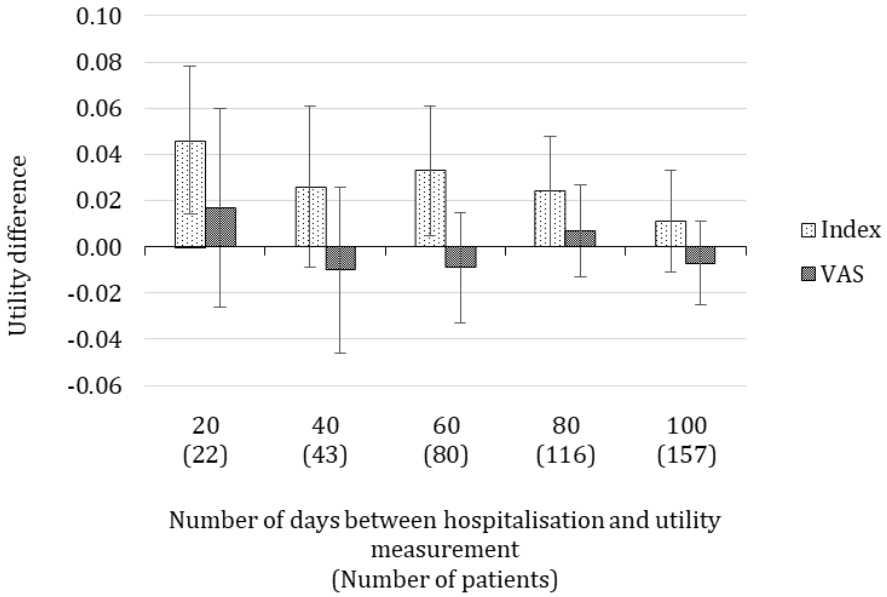


Figure 3.3 – Stratification of patients according to number of days between hospitalisation and utility measurement

Utility variation by primary admission cause

There were 456 records (78.22% of total number of hospitalisations) among the ten most common reasons for primary admission. The two causes with higher positive impact on the utility variation were respiratory/chest infection and ventricular tachycardia, whereas the causes with the highest negative impact on utilities were atrial fibrillation and myocardial infarction. Similarly to what was observed for the base case analysis, the calculated utility variations before and after hospitalisations for most of the primary admission causes registered high standard errors. The summarised results are presented in Table 3.2.

Table 3.2 – Utility variation by primary admission cause

Primary admission cause	**n (% total hospitalisations)**	**Utility variation (standard error)**	
		Index	VAS
Atrial Fibrillation	12 (2.06)	-0.102 (0.213)	0.010 (0.184)
Cardiovascular investigation	13 (2.23)	0.031 (0.046)	-0.125 (0.250)
Heart failure	226 (38.77)	0.014 (0.069)	-0.024 (0.060)
Myocardial Infarction	14 (2.40)	-0.123 (0.075)	-0.175 (0.181)

Other gastrointestinal cause	25 (4.29)	0.070 (0.173)	-0.018 (0.105)
Other not listed	87 (14.91)	0.041 (0.093)	0.018 (0.076)
Respiratory; Chest infection	22 (3.77)	0.106 (0.186)	-0.062 (0.085)
Stable Angina	13 (2.23)	0.032 (0.215)	-0.148 (0.212)
Unstable Angina	31 (5.32)	-0.068 (0.106)	0.008 (0.157)
Ventricular Tachycardia	13 (2.23)	0.300 (0.334)	0.185 (0.219)
Total	**456 (78.22)***		

* From a total number of 583 hospitalisations among sixty admission causes

Impact of the characteristics of the patient and of the hospitalisation in the variation in HRQoL

The descriptive statistics of the patient characteristics measured during the quarterly clinical visits and the results of the regression models used to assess the impact of these characteristics on the differences in utility pre and post hospitalisation are presented in Table 3.3. The model showed a statistically significant impact ($\alpha = 0.05$) of body mass index and serum creatinine in the index utility differences. For VAS utilities only serum creatinine was shown to explain the differences observed in statistically significant ($\alpha = 0.05$) manner.

Concerning the hospitalisation characteristics, none of these explanatory variables were shown to have a statistically significant ($\alpha = 0.05$) relationship with the variation in utility for both index and VAS utilities. The results for this model are summarised in Table 3.4.

Table 3.3 – Descriptive statistics of patient characteristics and their impact on utility pre and post hospitalisation

Variable	Descriptive statistics (mean [standard deviation])			Regression analysis Index			Regression analysis VAS		
	Pre hospitalisation	Post hospitalisation	Mean of Differences	Coefficient estimate	Standard error	p-value	Coefficient estimate	Standard error	p-value
Intercept	-	-	-	1.34×10^{-2}	2.39×10^{-2}	0.5748	2.12×10^{-2}	1.87×10^{-2}	0.259
Body mass index (kg/cm^2)	26.4 (5.9)	23.9 (9.9)	-2.5 (8.9)	1.73×10^{-2}	7.33×10^{-3}	0.0198*	7.45×10^{-4}	5.77×10^{-3}	0.897
Systolic blood pressure (mm Hg)	116.7 (21.3)	118.4 (22.6)	1.2 (21.5)	2.49×10^{-4}	1.39×10^{-3}	0.8579	-1.86×10^{-4}	1.07×10^{-3}	0.862
Diastolic blood pressure (mm Hg)	69.2 (10.9)	69.5 (11.5)	0.3 (14.2)	8.66×10^{-4}	2.24×10^{-3}	0.6991	2.01×10^{-4}	1.73×10^{-3}	0.908
Haemoglobin (g/dl)	8.8 (6.3)	8.8 (6.1)	-0.0 (2.5)	-1.77×10^{-3}	9.38×10^{-3}	0.8510	-1.10×10^{-3}	7.45×10^{-3}	0.882
Serum sodium (mmol/l)	134.6 (19.3)	133.4 (22.7)	-2.5 (26.5)	5.88×10^{-5}	1.44×10^{-3}	0.9674	7.60×10^{-4}	1.13×10^{-3}	0.503
Serum creatinine (µmol/l)	136.6 (64.4)	139.5 (58.3)	1.1 (55.2)	-9.70×10^{-4}	4.52×10^{-4}	0.0339*	-8.70×10^{-4}	3.55×10^{-4}	0.016*

* p-value < 0.05

Table 3.4 – Descriptive statistics of hospitalisation characteristics and their impact on utility variation pre and post hospitalisation

Variable	Descriptive statistics		Regression analysis Index			Regression analysis VAS		
	Mean	Standard deviation	Coefficient estimate	Standard error	p-value	Coefficient estimate	Standard error	p-value
Intercept	-	-	-9.11×10^{-3}	9.24×10^{-2}	0.922	9.93×10^{-2}	6.40×10^{-2}	0.122
# days before hospitalisation	52.7	39.0	-3.02×10^{-4}	7.78×10^{-4}	0.698	6.56×10^{-4}	5.42×10^{-4}	0.227
# days after hospitalisation	74.8	41.5	3.23×10^{-4}	7.67×10^{-4}	0.674	-9.15×10^{-4}	5.21×10^{-4}	0.081
Length of stay (days)	13.7	24.9	-1.30×10^{-3}	1.64×10^{-3}	0.429	-1.2×10^{-3}	1.23×10^{-3}	0.309
Intensive care (%)*	5.69	-	7.76×10^{-2}	1.01×10^{-1}	0.445	1.24×10^{-1}	7.63×10^{-2}	0.107

* dummy variable

Discussion

Quantifying the impact of hospitalisation on QoL is particularly relevant for informing cost-effectiveness studies designed to assess health technologies primarily aimed at reducing admissions, especially when compared to technologies aimed at reducing the decline of the patient health status. To the best of our knowledge, this is one of the first studies to address the impact of hospitalisation in HRQoL of heart failure patients. In view of readmission being a common event for these patients (147, 148) – with the first few weeks after discharge from hospital being the highest risk period; between 20-30% of patients are readmitted within 30 days, rising to 50% at 6 months (149) –, it seemed relevant to have an estimate of the impact of this event on the HRQoL of HF patients, thereby overcoming the use of utility decrement estimates based on the assumption that patients progress to the immediately worse NYHA class after hospitalisation (138). Using patient-level data we have calculated an empirical estimate for the difference between HRQoL before and after hospitalisation.

In this study we found a slight difference between the HRQoL measured before and after a hospitalisation: an increase in index utilities of 0.020 [95% CI: -0.020, 0.059] and a decrease of -0.012 [95% CI: -0.043, 0.020] for VAS utilities. Even though there is a discrepancy between the directions of this change, the small magnitude of the effect – further substantiated by the relatively large confidence interval around the mean and the similarity between the density curves of the two methods – indicates that there is no significant evidence of a difference between utility pre and post hospitalisation when using either of the utility elicitation methods. Nonetheless, there are two possible explanations for the difference between index and VAS utilities: (i) VAS utilities tend to be lower than index utilities for the same individuals (see Table 3.3); and (ii) changes in index utilities measured with the EQ-5D-3L are prone to "jumps", as they are only possible through a change in the patient self-assessment of his/her health state within the three possible levels – no problems, some problems, and extreme problems – for each of the five health dimensions. The five level EQ-5D (EQ-5D-5L) has since been introduced and it has proven to be a superior tool than EQ-5D-3L with respect to various measurement properties, enabling improvements in sensitivity and precision in health status measurement and the resulting utilities (150).

The findings for the base case analysis are further substantiated by the sensitivity analyses, except for the one that consisted of attributing 0 to the value of the utility post hospitalisation in patients who either died in hospital or before having an available measurement after the event. This analysis resulted in a decrease of -0.172 [95% CI: -0.222, 0.122] for the index utility and -0.133 [95% CI: -0.171, -0.096] for the VAS utility.

However, it is crucial to note that hospitalisations in heart failure patients are heterogeneous and, therefore, the impact of these hospitalisations on QoL is likely to depend on the underlying clinical cause for admission. For instance, a hospitalisation resulting from a temporary deterioration in the health of a patient, typical in revolving-door patients, may lead to an improvement in QoL measured before and after hospitalisation, whereas a stroke or other disabling event is likely to show the opposite. Further, it may be difficult to attribute hospitalisation to a single cause or to a single disease factor in a disease like heart failure, especially when considering all the comorbidities that are frequently associated with the disease. The small effect encountered for the base case analysis might be due to the offset of hospitalisations caused by different underlying problems in HF patients.

The results found when analysing the HRQoL pre and post hospitalisation by primary cause for hospital admission seem to suggest that it is possible to distinguish the impact on QoL for different types of hospitalisation (see Table 3.2). In that analysis, hospitalisations due to respiratory/chest infection and ventricular tachycardia showed an improvement in QoL when considering the index utilities measured before and after admission, while hospital admissions attributed to atrial fibrillation and myocardial infarction showed a negative variation in index utilities measured before and after admission. These results appear to be in line with the hypothesis postulated in the previous paragraph.

The regression analyses for explaining the observed variation in utilities before and after hospitalisation were inconclusive concerning the characteristics of the hospitalisation. However, the difference in body mass index (only for index utilities) and serum creatinine (both for index and VAS utilities) pre and post hospitalisation showed a significant effect on the utility variation, albeit no informed explanation for the mechanism of this effect can be provided, as it was not covered by the scope of this study.

Although international guidelines are clear in prioritising quality of life in the management of patients with HF (10, 12), their perception on their quality of life is not always prone to a straightforward assessment in a trial setting (151). Bosworth et al. (152) showed that psychosocial aspects and patient uncertainty about their prognosis are important components of quality of life among HF patients. Similarly, Heo et al. (153) found quality of life in patients with HF to be a multidimensional, subjective concept, affected by a variety of factors that do not only reflect HF symptoms and limitations in their daily life due to those symptoms, but also their active pursuit of happiness and relationships with others. Other factors such as anxiety, general distress, or depression have been shown to decrease QoL amongst HF patients (25, 154), whilst interventions aimed at improving patient self-care proved to have positive impact on QoL (155). Following on these thoughts, it can be argued that

hospitalisation is a source of distress for patients, who would therefore experience a decrease in HRQoL. In fact, a study by Harrison et al. (156) showed significant improvements in HRQoL associated with lesser use of emergency rooms – even though one can also argue on inverse causality, i.e. that fewer visits to the emergency rooms may be due to better health and thus higher HRQoL. Another study, by Lewis et al. (137), found that myocardial infarction survivors experiencing a nonfatal cardiovascular event (hospitalisation for heart failure, recurrent myocardial infarction, stroke, or sudden death/cardiac arrest) had a significant worsening of their HRQoL when compared to the ones who did not experience such event, suggesting that reducing nonfatal cardiovascular events might affect longitudinal changes in HRQoL.

Having an accurate estimate of the utility variation attributed to hospitalisation in HF patients would be a great addition to the economic evaluation arsenal. In fact, there are discrete event simulation models published in the cost-effectiveness literature that use "hospitalisation" as an "event" (157). Especially for these cases, a good estimate of the (dis)utility of a hospitalisation would be of great value.

Limitations

The variation between utilities pre and post hospitalisation showed a different magnitude from what was hypothesised and the value found for that variation was surrounded by a lot of uncertainty. There are some possible explanations that can be identified in the scope of the limitations of this study.

First, the patient population included in the analysis was already in a very advanced stage of the disease: (1) coping with chronic disease has been described to have a positive influence in the QoL perception of the ill patient (158) and (2) the fact that these patients have been previously hospitalised – as this was one of the inclusion criteria of the trial – may desensitise them to subsequent hospitalisations. Secondly, the measurement of HRQoL is not done at a particular moment related to the hospitalisation; in order to be comparable, the utility measurement should be done at admission, discharge, or, preferably, both – the mean number of days before the HRQoL measurement before hospitalisation is 52.7 and after hospitalisation is 74.8; both with large standard deviations (cf.

Table 3.4). The results of the third sensitivity analysis should also be discussed in this context: in spite of the lack of statistical significance, results in Figure 3 seem to suggest that the magnitude of the utility difference is higher when the HRQoL is measured closer to the hospitalisation date. Thirdly, we did not have information that would allow for adjusting for other factors that might have affected changes in HRQoL, including changes in medications and/or any surgical procedures done during hospitalisation. And finally, the EQ-5D-3L is an utility measurement tool that assesses global health status and that may not be as responsive as a disease-specific instrument like the Minnesota living with heart failure questionnaire (MLHFQ) (159).

Recommendations for future research

Paying attention to the main issues that have been discussed so far, a few points should be stressed.

Some standardisation regarding the moments at which HRQoL is measured is desirable. This concept should be applicable not only in a controlled setting but also in current clinical practice. Conducting EQ-5D questionnaires or using other tools for measuring utilities may generate data that could turn out to be important in the development of guidelines for the management of heart failure. Special attention should be paid to the variance observed in HRQoL from clinical trials and the clinical practice. Some possible explanations for this variance are: (1) diseases with multiple comorbidities, where the trial population is often not representative of the real patient population, and (2) the Hawthorne effect, i.e. the mere attention paid by clinical trial personnel to study subjects, which may have beneficial effects on the QoL of participating trial patients (160, 161).

In the particular case of home telemonitoring – from which the population in this study originated – daily measurements of HRQoL could be performed. Considering most telemonitoring settings it is not expected that these measurements would constitute an increased burden for patients. Yet, the generation of longitudinal utility data would allow for investigating QoL as a predictor for hospitalisation. Health deterioration could be captured by trends in the data regarding patient-reported HRQoL. The analysis of these data could potentially result in the development of clinical decision rules or diagnostic algorithms that could avoid unnecessary hospitalisations, leading to potential cost savings and better health outcomes in the management of heart failure.

Considering that disease-specific instruments for HF (e.g., MLHFQ) can be more informative on patient perceived health status, a formal link between the outcomes of these questionnaires and a measure of utility should be established. In this way researchers could have access to more accurate information on patient-reported

outcomes without compromising utility measurements that are normally used for economic evaluations.

And finally, HF-related research should focus on the determinants of HRQoL in heart failure patients. Although NYHA is a widespread classification of the severity of HF symptoms, the current capabilities for data collection and data generation are immense. They should be explored in order to open up possibilities for new classifications that could better suit the need of efficient management of heart failure patients.

Conclusions

Knowing the impact of hospitalisation on health-related quality of life is particularly relevant for informing cost-effectiveness studies designed to assess health technologies aimed at reducing hospital admissions. Through using patient-level data it was possible to estimate the variation in utilities before and after the average hospitalisation and for hospitalisations due to the most common causes for hospital admission. These estimates for (dis)utility could be used in the calculations of effectiveness on economic evaluations, especially when discrete event simulations are the employed modelling technique.

Chapter 4

An ensemble algorithm to predict clinical deterioration in heart failure patients using a telehealth programme

Background

Heart failure (HF) is the number one cause for hospitalisation in adults over 65 years of age in the United States (US) (162). Hospitalisation in heart failure is a consequence of acute (decompensated) heart failure [A(D)HF], a clinical condition characterised by a rapid onset or worsening of symptoms and/or signs of HF that requires urgent evaluation and treatment, typically leading to an urgent hospital admission (10, 12). Every year more than 1 million patients are hospitalized with a primary diagnosis of heart failure, costing Medicare over $17 billion (163). The readmission rate for these patients is considerable, with 8.4% of patients being readmitted within 30 days of discharge, 13.4% within 60 days, and 16.7% within 90 days (164). In recent years major cardiology associations have been calling for the development of better prediction techniques that could lead to earlier diagnosis and treatment, thus improving health outcomes of ADHF (165, 166).

Current HF management programmes are based on active surveillance of signs and symptoms of ADHF, along with medication-related and disease-specific education to optimise treatment. Alternative models relying on information communication technology for self-monitoring and for remote control signs and symptoms control via advanced telemonitoring technology (e.g., electronic transfer of physiological data measured by the patient) have been tested recently (167, 168). The data generated by these technologies permits their use as predictors of specific outcomes, which can then be used for developing predictive models assisting medical personnel in their decision-making process (65-68). After implementation and testing in a clinical setting, those models can be translated into clinical decision rules for clinical practice (69). However, some of these rules lack rigorous assessment of their predictive accuracy and/or their potential to improve current decision-making (169).

In the specific case of home telemonitoring (HTM) for HF, evidence shows that data-driven approaches looking at trends and patterns of recorded parameters change seems to improve the accuracy of detecting disease deterioration when comparing to clinical decision rules (170-173). For that reason, and because a large number of parameters can be measured with HTM, it is expected that advanced algorithms with better performance measures (e.g., sensitivity and specificity) will likely result in time efficiencies and improved clinical decision making through the generation of alerts in a manner that is intuitive and that can be used by clinicians with a higher degree of confidence (174). Diagnostic algorithms are mathematical relationships that use a wide range of collected data for outputting the likelihood of an event (e.g. hospitalisation or death) and use this output for prioritising patients with regards to their treatment or, alternatively, through raising an action-triggering alarm if the probability of having the event exceeds a pre-defined threshold.

This chapter describes the concept of an ensemble algorithm for estimating the risk of HF-related hospital admissions using an ambulatory telehealth programme combining clinical software and in-home remote monitoring technology. The algorithm intends to provide clinical professionals with a global risk score and to use this score to help defining follow-up actions based on evidence-based thresholds.

Methods

Telehealth programme (intervention)

The telehealth programme (henceforth referred to as intervention) is a remote monitoring technology that measures vital signs and inputs from symptom surveys (shortness of breath, oedema, and fatigue), combined with a clinical software available for data monitoring and assessment by a clinical team. All measurements and surveys are planned to be filled in daily by the patients and remotely transmitted to the clinical software installed in the health care provider facilities.

Data

Sites

The data for the development of the algorithm were collected from two large non-profit health care systems in the US where the intervention was implemented, which generated two distinct datasets containing the same elements.

Data elements: vitals

The data on vitals comprise daily measurements taken during the morning of body weight, systolic blood pressure, diastolic blood pressure, heart rate, and blood oxygen saturation. Each raw measurement record includes a patient identification and a time stamp.

Data elements: surveys

The data on surveys consists of a score derived from the answers given by patients to multiple choice questionnaires on oedema, fatigue, shortness of breath, and activity status administered remotely. Similarly to the vitals data, they were collected daily.

Data elements: outcomes

Outcome data were derived from electronic medical records of the sites from which data were collected. There were no data on emergency department visits or deaths; readmissions were the only type of outcome for which data were available.

The feature set

The feature set consists of the data – raw or transformed – that are used as inputs for the models that constitute the algorithm. The following raw data were considered: (i) age (years), (ii) gender (male/female), (iii) date of enrolment for current telemonitoring engagement, (iv) living arrangement (alone/with a partner), (v) number of past telemonitoring engagements, (vi) patient diagnoses, (vii) systolic blood pressure (mmHg), (viii) diastolic blood pressure (mmHg), (ix) body weight

(pounds), (x) heart rate (beats per minute), (xi) peripheral capillary oxygen saturation (percentage), (xii) answer to oedema question of symptom survey, (xiii) answer to fatigue question of symptom survey, (xiv) answer to shortness of breath question of symptom survey, and (xv) answer to activity question of symptom survey. Some of these raw data elements were directly used as features in the algorithm, while other features used in the algorithm were calculated by transforming the raw data with a wide range of arithmetic operations (mean and standard deviation for determined periods of time, deviation of measurement from baseline value, trend scores calculated through linear regression, etc.).

Design and development of the algorithm

Prediction window

The algorithm intends to use measurement data from the intervention and turn them into accurate alarms to flag patients in need of further attention from a healthcare provider. It uses the full data on the feature set and it calculates a risk score for being hospitalised within determined number of days called the prediction window. For instance, a 30-day prediction window means that the risk score predicted by the algorithm is the probability of a hospitalisation happening in the upcoming thirty days. During the development of the algorithm, various prediction windows were tested.

Solo versus ensemble algorithm

A solo algorithm uses the data in the feature set for calculating risk scores based on single regression models (e.g., logistic regression), while ensemble algorithms combine the outcome of different predictive models (layer0 models) into a final risk score. The latter approach provides a clear advantage when the models are fed with different feature subsets, when the models are based on different regression techniques, or both.

The ensemble algorithm presented in this chapter results from dividing the feature vector into different subsets (“surveys”, “vitals”, “advanced”, and others) and from training the model per feature subset using the least absolute shrinkage and selection operator (LASSO) technique (175). The LASSO technique is used for improving the prediction accuracy and interpretability of regression models by altering their fitting process in order to select only a subset of the provided covariates for being used in the final model. LASSO was applied to each layer0 model in order to prevent overfitting and to eliminate those features which were not informative in predicting patient deterioration and hospitalisation.

Additionally, the ensemble algorithm can further introduce another layer of prediction between the output of layer0 models and the final score – the layer1 model (referred to as a meta-learner) (176).

Figure 4.1 presents a graphical overview of the architecture of the solo and the ensemble algorithms.

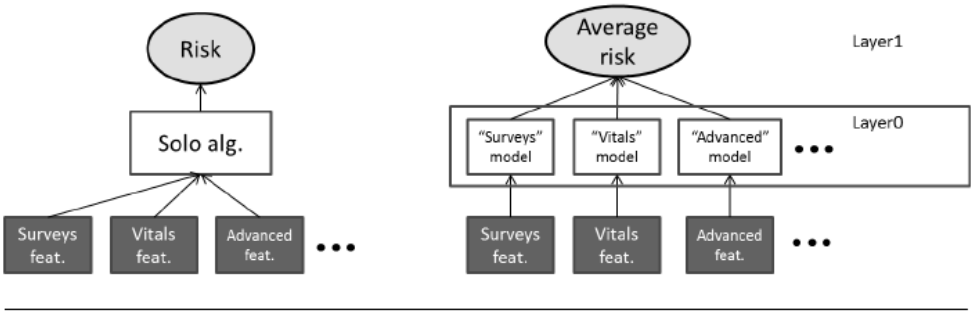


Figure 4.1 – Graphical overview of the architecture of a solo (left) and ensemble (right) algorithms

Comparison to a Standard Protocol

At the time of the algorithm development there was an available set of decision rules (henceforth referred to as the Standard Protocol) from one of the sites used in the data collection. The Standard Protocol used blood pressure, pulse, weight, and blood oxygen saturation measurements for generating a risk profile (low, medium, or high) for each patient on any given observation day. A comparison between the sensitivity, specificity, and positive predictive value (PPV) of the Standard Protocol and the algorithm can be done at the three operating points of the receiver operating characteristic (ROC) curve of the algorithm at which the specificity is the same as the one for the three risk profiles with the Standard Protocol (0.90, 0.76, and 0.53 for high, medium, and low risk, respectively).

Therapeutic threshold

The main objective of the algorithm is to give an early indication (e.g., raising an alarm) indicating whether the patients should be screened, or, in other words, flag when the risk score calculated by the algorithm exceeds a pre-defined risk threshold for acting clinically. As such, there is a need to rationally define the threshold at which action should be taken. This issue can be tackled with the therapeutic threshold concept (177).

The therapeutic threshold concept starts with a decision between administering a treatment and withholding it: in our case, screening or not screening based on an alarm raised from the algorithm output. Regardless of this decision, let us take any

given patient who has an unknown probability p of having the disease and a probability of $1 - p$ of not having it: in our case, it is represented by the ground truth of being or not being hospitalised. Thus, considering the decision to screen as the treatment and the probability of being hospitalised as the disease, there are four possible outcomes: (a) screening + hospitalisation; (b) screening + no hospitalisation; (c) no screening + hospitalisation; and (d) no screening + no hospitalisation.

For each of the four possible outcomes, we can assign a value (V) in any relevant unit (utility, remaining life-years, monetary units, etc.). The expected value for outcome (a) can thus be represented by $EV_a = p.V_a$. Similarly, the expected value of screening can be represented by $EV_{screen} = p.V_a + (1 - p).V_b$, while the expected value of not screening would be represented by $EV_{no\ screen} = p.V_c + (1 - p)V_d$. Applying these principles, one can decide whether to screen or not to screen by choosing the course of action with the higher expected value. The indifference point for treating is found when $EV_{screen} = EV_{no\ screen}$. Working on this equation and solving it for p leads to $p = \frac{(V_d - V_b)}{(V_a - V_c + V_d - V_b)}$, where p is the probability of hospitalisation at the indifference point, or in other words, the threshold for the hospitalisation probability at which it is indifferent to screen or not to screen.

In the particular case of the analysis of the therapeutic threshold of the algorithm, monetary units were used for V. Assuming that screening correctly identifies the patients who would have been hospitalised, it seems reasonable to rename the outcomes as follows: (a) planned hospitalisation, (b) screening and no hospitalisation, (c) unplanned hospitalisation, and (d) no screening and no hospitalisation (see Figure 4.2 for the schematic representation of the corresponding decision tree for the therapeutic threshold using the algorithm).

For calculating the value of planned and unplanned hospitalisations, the length of stay in hospital was used as a proxy. We included all studies identified in the review by Inglis et al. (178) where the mean number of days in hospital and number of hospitalisations for both home telemonitoring (planned hospitalisation) and usual care (unplanned hospitalisation) were reported or could be calculated (126, 179-186). The length of stay average was subsequently calculated for both treatment alternatives and weighted it with the number of hospitalisations recorded in each study. These values were multiplied by the daily cost of hospitalisation for Medicare beneficiaries with heart failure (adding the costs of screening for the value of the planned hospitalisation) (164). The costs of screening consist of the average costs reported by Medicare facilities for the Healthcare Common Procedure Coding System code G0406, which correspond to a 15-minute follow-up telehealth consultation (187, 188). The value of the outcome not screening and no hospitalisation is considered to be $0.

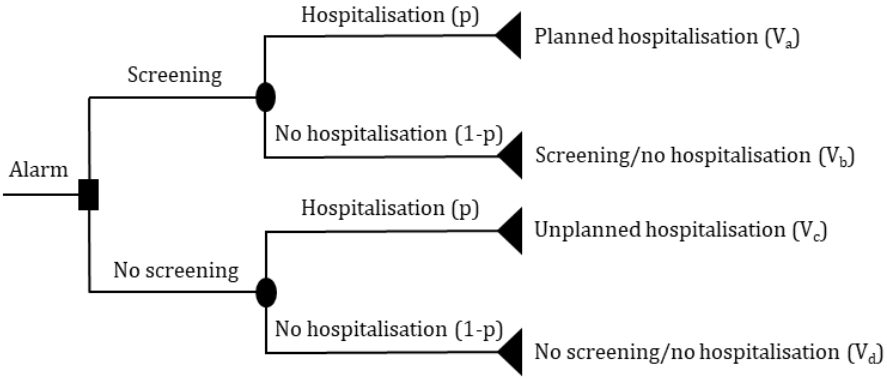


Figure 4.2 – Decision tree when facing an alarm triggered by the algorithm

Results

Design and development of the algorithm

The final base case of the ensemble algorithm consists of twenty layer0 models and a prediction window of 5 days. Given the restricted availability of data, the outputs of layer0 models are aggregated without any layer1 model. The final risk score is therefore the simple average of all outputs from layer0 models, i.e. the average of the risk scores predicted by the different logistic regressions.

The average area under the receiver operating characteristic (AuROC) for the ensemble algorithm is 0.81 (± 0.01 standard error) for a prediction window of 5 days. Table 4.1 shows the predictive performance of the ensemble model varying the length of the prediction window.

Table 4.1 – Predictive performance of the ensemble algorithm with different prediction window length

Window length (days)	Mean AuROC	Standard error
1	0.82	0.01
2	0.82	0.02
3	0.81	0.01
5	0.81	0.01
10	0.78	0.02

The predictive performance of the Standard Protocol is characterised by an AuROC of 0.55. The comparison between the sensitivity, specificity, and positive predictive value (PPV) of the algorithm and the Standard Protocol at the three operating points of the Standard Protocol (low, medium, and high risk) as reference points is shown in Table 4.2.

Table 4.2 – Comparison of the sensitivity, specificity, and PPV between the algorithm and the Standard Protocol

Risk	Sensitivity		Specificity		PPV	
	Algorithm	*Standard Protocol*	*Algorithm*	*Standard Protocol*	*Algorithm*	*Standard Protocol*
High	0.39	0.19	0.90	0.90	0.024	0.020
Medium	0.67	0.35	0.76	0.76	0.018	0.016
Low	0.91	0.52	0.53	0.53	0.013	0.012

Therapeutic threshold

The average daily costs of HF-related hospitalisation for Medicare beneficiaries with heart failure found were $2,069.45, and the calculated average weighted length of stay in hospital was 7.72 days for home telemonitoring and 7.89 days for usual care (see Table 4.3), resulting in total costs for a planned hospitalisation (V_a) of $16,023.55 and for an unplanned hospitalisation (V_b) of $16,324.87. The estimated average costs for screening (i.e. a 15-minute follow-up telehealth consultation; V_c) were estimated at $39.97 and the cost of not screening (V_d), as per the described methodology, was deemed $0. Through applying these values to the indifference point equation, using costs as the outcome, one reaches a probability of indifference of 0.117, or, in other words, a patient should be screened if his/her probability of being hospitalised as calculated by the algorithm is higher than 11.7%.

Table 4.3 – Calculation of the weighted average days in hospital with and without treatment

Study	Average days in hospital (home telemonitoring)	Average days in hospital (usual care)	# hospitalisations (home telemonitoring)	# hospitalisations (usual care)	Weighted average days in hospital (home telemonitoring)	Weighted average days in hospital (usual care)
Balk 2008 et al. (179)	7.4	7.9	103	96	1.80	1.79
Dendale 2012 et al. (180)	2.5	4.6	64*	65.6*	0.38	0.71
Koehler 2011 et al. (181)	5.3	4.9	31	36	0.39	0.42
Lyngå 2012 et al. (182)	7.5	7.7	70	70	1.24	1.28
Soran 2008 et al. (183)	10.0	9.3	29	36	0.68	0.79
Vuorinen 2014 et al. (184)	0.7	1.4	8	13	0.01	0.04
Woodend 2008 et al. (185)	2.7	3.8	28.52*	28.91*	0.18	0.26
Cleland 2005 et al. (126)	14.5	15.5	67	33	2.29	1.21
Mortara 2009 et al. (186)	13.5	13.3	24	44	0.76	1.38
Average	**7.12**	**7.59**			**7.72**	**7.89**
Difference	**-0.47**				**-0.16**	

*decimal numbers result from the # hospitalisation being calculated using the reported rate for hospitalisation

Discussion

This chapter aimed at describing an ensemble algorithm to estimate the risk of HF-related hospital admissions using data generated with a home telemonitoring system and to determine the therapeutic threshold for raising an alarm for prompting the screening of the monitored patients. There is already literature published about risk-prediction models for ADHF, both in the form of clinical prediction scores (189, 190) and algorithms (170-173, 191).

The dissemination of telemonitoring technologies encompasses the generation of large amounts of data that will inevitably confront health professionals with the best way of using them in order to make correct clinical decisions, creating a problem that can be referred to as information overload. To overcome this issue and manage healthcare information effectively, there is a need to adapt existing systems for filtering information or to create ones that respond to the challenges of healthcare communication in the digital age (192). The activities carried out during the algorithm development allow for shedding some light on the relevance of collecting telemonitoring data on a daily basis to predict adverse events. It became apparent that in order to build a reliable predictive model there is a need for collecting high quality data as well as recording reliable information on the ground truth data points, especially with regards to the type and date of the event to be predicted. On this matter, the development of an ensemble algorithm proves a valid alternative for overcoming the aforementioned issues. Still on the topic of data generation and data collection, it is worth mentioning that a validated algorithm for a technology such as telemonitoring, where data are constantly generated and can be fed to the model for recalibration, has a great chance of continuously improving its predictive accuracy through machine learning.

The ensemble algorithm presented allows for overcoming the main constraint found in using a simple logistic regression model in the available dataset, i.e. the need to feed the model with a complete feature vector in order to generate a risk score. More specifically, it dealt with the problem regarding the level of completeness of the data, both at an observation level (e.g., any day in which the patient does not measure a vital sign) and at a patient level (e.g., the protocol of a certain patient not including the measurement of a particular vital sign for the entire monitoring period), which resulted in a highly reduced number of complete observation days.

The algorithm shows improved predictive performance when compared to the Standard Protocol, which suggests that using the algorithm can improve the decisions made by clinical teams in order to avoid hospitalisation of patients with HF. The predictive parameters of the algorithm, however, depend on the operating point

chosen as well as the prediction window used in the analysis. As expected, the mean AuROC increases as the prediction window's length decreases. This can be explained by a shorter prediction window only labelling the days closer to the event as positives (i.e. observation days when deterioration occurs), making the signs of deterioration more pronounced. However, it should be borne in mind that the final choice in terms of window length is driven also by the prospective predictive performance of the model and the possibility of taking action in case of an alarm – thus the 5-day prediction window.

The therapeutic threshold analysis aimed at defining a threshold for raising an alarm based on the monetary values attributed to the outcomes deriving of screening the patients and whether or not these same patients would be hospitalised. Although this can provide an evidence-based value for the threshold, this approach would also estimate different thresholds if one would use other units for the values of the outcomes. The full value of the algorithm should be assessed by taking into account that different operating points of the algorithm lead to different rates of true and false positives and true and false negatives. Thus, the right operating point of the algorithm is subject to the costs and benefits of the false positives and false negatives, and the trade-off between these two groups. Since the daily calculation of the outcome of the algorithm can be considered a diagnostic test, it should be properly assessed through a full cost-effectiveness analysis in which all costs and outcomes of using the algorithm should be included.

Limitations

One limitation of the presented algorithm relates to using hospital readmission as the only outcome measure. The data used for training the algorithm could be extended to other outcomes that result from heart failure deteriorations, such as emergency department visits and death. Still related to the scope of the presented algorithm, the ensemble algorithm provides a risk score reflecting the likelihood of being readmitted to the hospital within 5 days for HF-related reasons and it does not consider the broader spectrum of all-cause readmissions.

Another limitation derives from not including any information concerning interventions performed while patients were telemonitored in the predictive model (medication changes, modifications of the diet, activity level changes, etc.). All of these could have prevented adverse events, thereby leading to changes in the patients' deterioration pattern that cannot be tracked by the models included in the algorithm, as none of the features used in the dataset was able to provide information on this matter.

The algorithm was developed on HF-patient data and should be applied on patients with HF who are monitored with the technology under analysis. Performance of the algorithm on the other patient groups who are eligible to use telemonitoring technologies (e.g., COPD patients) cannot be guaranteed.

Recommendations for future research

Future research on this topic should be aimed at investigating other types of feature data, mainly body mass index, ethnicity, socio-demographic, and economic factors. Additionally, there should be a particular focus on disease-specific parameters that have been shown to have an impact on hospitalisation for patients with HF (193), such as left ventricular ejection fraction (194) and NT-proBNP (195, 196).

One key aspect of the implementation of predictive modelling is how the model output is embedded in the work flow of the professionals and how clinical outcomes can be improved through the use of the model. In other words, the model should support decisions through making use of the available information and providing a concise description of the clinical status of the patient. As such, the risk score should not be an additional source of information that could hinder decision-making, but rather an easy interpretable tool that provides clear indication on follow-up actions to the health professional. In this sense, different solutions could be investigated. For instance, raising a red flag for deteriorating patients and displaying their risk score of the last few days, sorting patients according to their risk score by means of prioritising clinical intervention, or creating a tiered system for alarms (e.g., low, medium, and high risk). Accordingly, the involvement of skilled professionals in the design of the user interface for the algorithm is highly desirable, as clinical processes and goals of different health care providers may differ (197).

Conclusions

The ensemble algorithm appears to be a tool with high potential for improving the clinical decision-making and reducing the readmission of patients with HF under home telemonitoring programmes. The predictive value of the algorithm was demonstrated by the significant increase in the AuROC when compared to the Standard Protocol.

The therapeutic threshold analysis determined that the threshold for raising an alarm should be set to when the output calculated by the algorithm is higher than 0.117. However, the added value of the algorithm in the clinical setting should be investigated through a full cost-effectiveness analysis that could properly assess the costs of implementing and using the technology versus the anticipated effectiveness resulting from the early detection of disease deterioration and the subsequent reduction of hospitalisations.

Chapter 5

Modelling early warning systems: construction and validation of a discrete event simulation model for heart failure

Fernando Albuquerque de Almeida, Isaac Corro Ramos, Maureen Rutten-van Mölken, and Maiwenn Al

Value in Health (In Press)

Abstract

Objectives: Developing and validating a discrete event simulation model that is able to model heart failure patients managed with usual care or an early warning system (with or without a diagnostic algorithm) and to account for the impact of individual patient characteristics in their health outcomes.

Methods: The model was developed using patient-level data from the TEN-HMS study. It was coded using R and validated along the lines of the Assessment of the Validation Status of Health-Economic decision models tool (AdViSHE). The model includes 20 patient and disease characteristics and generates 8 different outcomes. Model outcomes were generated for the base-case analysis and used in the model validation.

Results: Patients managed with the early warning system, when compared to usual care, experienced an average increase of 2.99 outpatient visits and a decrease of 0.02 hospitalisations per year, with a gain of 0.81 life years (0.45 quality-adjusted life years) and increased average total costs of 11,249€. Adding a diagnostic algorithm to the early warning system resulted in a 0.92 life year gain (0.57 quality-adjusted life years) and increased average costs of 9,680€. These patients experienced a decrease of 0.02 outpatient visits and 0.65 hospitalisations per year, while avoided being hospitalised 0.93 times. The model showed robustness and validity of generated outcomes when comparing them to other models addressing the same problem and to external data.

Conclusions: This study developed and validated a unique patient-level simulation model that can be used for simulating a wide range of outcomes for different patient subgroups and treatment scenarios. It provides useful information for guiding research and for developing new treatment options by showing the hypothetical impact of these interventions on a large number of important heart failure outcomes.

Introduction

Decision analytical models (henceforth models) are key instruments in the toolbox of health economists. Models are the resource by which researchers represent the complex reality in a more simplistic and comprehensible manner or by which experiments that are infeasible or impracticable are simulated (93). In the health economic context, through exploring hypothetical scenarios and alternative treatment strategies in order to identify the most efficient allocation of healthcare resources, models are used to inform decisions when significant real-world data are not available (198).

Heart failure (HF) is a complex clinical syndrome that results from any structural or functional impairment of ventricular filling or ejection of blood (10, 11). HF is characterised by typical symptoms such as breathlessness, ankle swelling, and fatigue and signs like elevated jugular venous pressure, pulmonary crackles and peripheral oedema (12). The main disease severity indicator used to describe HF is based on measurement of the left ventricular ejection fraction, which results in a distinction between HF with preserved, mid-range, and reduced ejection fraction – each with different underlying aetiologies, demographics, co-morbidities and response to therapies (13). The New York Heart Association (NYHA) functional classification is an alternative classification system that is used to describe the severity of symptoms and exercise intolerance, providing useful and complementary information about the presence and severity of the disease, and thus guiding patient pathways in HF treatment (14). HF is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. From a medical perspective, the goals of managing patients with HF consist in improving their clinical status, functional capacity, and quality of life, preventing hospital admissions and reducing mortality (39-41).Early warning systems (EWS) in the context of health care are timely surveillance systems that collect clinical information in order to anticipate health deterioration and trigger prompt intervention, thus improving prognosis and treatment outcomes (58). Broadly speaking, EWS consist of three main elements: (i) monitoring and collection of clinical data (e.g. vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a singular combination of signs and symptoms – that trigger an alarm and follow-up actions (59).

Diagnostic algorithms (DA) are predictive mathematical relationships that use a wide range of data collected by EWS for calculating the likelihood of an event (e.g.

hospitalisation or death). These algorithms are used for assisting medical personnel in their decision-making process (65-68) by translating their output into clinical decision rules for clinical practice, for instance, by prioritising patients according to their likelihood of having an event or by raising an action-triggering alarm if the probability of having that event exceeds a pre-defined threshold (69).

A previous systematic literature review of models used in the economic evaluation of EWS for the management of HF patients found that all published models were either decision trees or Markov models (59). However, due to the specific features of EWS in the context of HF, the flexibility for modelling complex systems provided by discrete event simulation (DES) models makes them arguably better option for the assessment of the (cost)-effectiveness of EWS (133, 199-202). DES or patient-level models (both terms will be used interchangeably henceforth) are a type of model that has been increasingly used in the health economics field, not only because of the advances in computing technology and dedicated software but also because of their flexibility and potential for modelling complex diseases (199-201, 203). One of the main advantages of DES modelling is the ability to use individual patient characteristics as explanatory variables for predicting disease pathways of simulated patients. In order to compare the cost-effectiveness of treatment strategies targeted at changing individual patient characteristics, DES models accounting for those characteristics and outputting a wide variety of (intermediate) outcomes are desirable. However, in order to be useful tools for decision-making regarding the problem at hand, DES models must accurately reflect disease pathways and their management (204).

The two main objectives of this study were: (i) developing a DES modelling framework for patients with HF managed with EWS – with and without a DA – that is able to model patients across the whole treatment pathway until death, taking into account the evolution and impact of individual patient characteristics in the outcomes of each individual patient, and (ii) justifying the model structure chosen and validating the model through the use of the AdViSHE questionnaire and the model outcomes generated in the base-case analysis.

Methods

Starting population of the model

The starting population of the model consisted of the patients who participated in the Trans-European Network – Home-Care Management System (TEN-HMS) study (126). This trial investigated the impact of using home telemonitoring (HTM; n = 168), nurse telephone support (NTS; n = 173), and usual care (UC; n = 85) in hospital admissions, hospital days, and rates of mortality. Patient-level data from the trial were used in the construction and validation of the model.

The simulated model population consisted of a set of randomly drawn patients (with replacement) from the database containing the patient-level data of the starting population. Table 5.1 shows the baseline characteristics of the starting population and of the simulated model population for 1,000 patients.

Table 5.1 – Patient and disease characteristics of the starting population and of the simulated model population of 1,000 patients

	Baseline characteristics of the starting population (TEN-HMS study)	**Simulated model population for 1,000 patients**
Sample size	426	1,000
Ejection fraction (EF), % (mean)	25.06	24.86
Age, years (mean)	67.56	67.76
Systolic blood pressure (SBP), mm Hg (mean)	114.24	114.53
Body mass index (BMI), kg/m2 (mean)	26.17	25.94
Creatinine, µmol/l (mean)	135.71	136.49
NYHA class 1, %	18.5	17.5
NYHA class 2, %	43.4	42.8
NYHA class 3, %	31.0	33.3
NYHA class 4, %	7.1	6.4
Gender (male), %	77.5	75.8
Smoker, %	12.2	11.9
Diabetes, %	35.0	37.3
Chronic obstructive pulmonary disease (COPD), %	24.4	21.2
Recent diagnosis, %	43.9	41.8

No beta-blocker medication, %	37.3	36.7
No ACE inhibitor medication, %	18.5	17.5
Myocardial infarction, %	56.8	56.2
Chronic atrial fibrillation, %	26.3	27.8

Abbreviations: ACE, angiotensin-converting enzyme; BMI, body mass index; COPD, chronic obstructive pulmonary disease; EF, ejection fraction; NYHA, New York Heart Association; SBP, systolic blood pressure.

Three interventions were considered in the model: (i) UC – patient management plan implemented by the patient's primary care physician, (ii) EWS (EWS without a DA) – proxied by HTM (described in detail in the TEN-HMS original publication (126)), and (iii) EWS+DA (EWS with a DA) – intervention (ii) with the addition of a DA (described in the following section).

Conceptualisation of early warning systems and the diagnostic algorithm for the management of heart failure

We conceptualised the EWS and the DA for the management of HF in the model from a clinical perspective, i.e. we have not simulated their impact in the actual pathogenetic process of the disease but rather how they manifest in clinical practice through their impact on each of the events considered in the model. In the scope of HF, the EWS collects clinical information such as vital signs, biomarkers, inputs from surveys, etc. – daily in our case – and uses it for changing the chance of death and/or hospitalisation. The effect of the EWS is captured by the difference of time-to- hospitalisation and time-to-death of HTM (the EWS in the context of our analysis) when compared to UC. The additional effect of the DA is captured by the possibility of avoiding hospitalisations as described in the following paragraphs.

In our instance, the DA is a mathematical feature that uses clinical data for calculating the likelihood of hospitalisation and raises an action-triggering alarm if the probability of being hospitalised exceeds a pre-defined threshold. It is added to the EWS as a way of automatically analysing the collected data in the EWS. In this framework, we can interpret the alarm as a diagnostic test: if an alarm is raised, the test is positive; if not, the test is negative. We can then consider the event of interest (hospitalisation) as "having disease/condition" and not being hospitalised as "not having disease/condition".

The interpretation of the statistical measures of the performance of a binary classification test in the context of the model can be described as follows: (i) when the simulated event is a hospitalisation, the sensitivity represents the probability of correctly detecting that hospitalisation. The final probability of avoiding a

hospitalisation can be achieved by multiplying the sensitivity of the test by the probability of avoiding a hospitalisation in the case of having correctly predicted it (e.g. assuming the sensitivity of the alarm is 0.8 and that 80% of the correctly predicted admissions can be avoided, then 0.8 x 80% = 64% is the overall probability of avoiding a hospitalisation). (ii) Regardless of the simulated event, there are as many diagnostic tests (alarm/no alarm) as there were days elapsed between the previous event and the current one. The model calculates the number of false positives (alarms for which there were no hospitalisation) in that period through multiplying the number of elapsed days by the false positive rate (FPR) of the DA (e.g. if there were 45 days between the previous and the current events and the FPR of the DA is 0.40, there were 18 false alarms during the period between both events).

Model structure

The main elements of the model are entities, attributes, events, procedures, outcomes, and relationships. The entity is the modelling representation of the patient (hereafter treated in the masculine form). Attributes are the characteristics of that patient, which can either be fixed throughout the simulation (e.g. previous history of myocardial infarction) or change over time (e.g. age). Events are relevant moments in the simulation that are recorded for reconstructing the clinical history of the entity; the model determines which event will happen next by calculating the lowest time-to-event of competing events. Procedures are the means by which the model processes events, following a decision-analytical logic that simulates the clinical pathway of the entity. During each procedure, attributes of the entity are re-evaluated, updated, and outcomes are generated and recorded. Outcomes are the elements that aggregate the information generated by the model and that allow for drawing conclusions from the performed simulations. Relationships are the model elements that link entities, attributes, events, procedures, and outcomes together through mathematical and/or logical terms defined in the model's code.

For ease of description of the model flow, elements are enclosed within <>, each with a subscript, depending on the type of element we are referring to (Ent, entity; A, attribute; E, event; Proc, procedure; O, outcome). At the start of simulation, a <patient>$_{Ent}$ is randomly drawn (with replacement) from the database containing the patient-level data of the starting population (patients participating in the TEN-HMS trial). Attributes are assigned to <patient>$_{Ent}$ based on the patient characteristics found at baseline in the dataset and calculates the time-to-event for each of the following competing events: <outpatient.visit>$_{E}$, <hospitalisation>$_{E}$, and <death>$_{E}$. Time-to-event depends on the individual attributes of the <patient>$_{Ent}$ at the time of the simulation. The lowest time-to-event determines which event will be processed next. The event is renamed as a procedure and a decision-analytical logic for each of

the different procedures determines the pathway of the patient. In <outpatient.visit>$_{Proc}$, time, costs, life years, and quality-adjusted life years (QALYs) are recorded, the selected attributes are updated, and the updated <patient>$_{Ent}$ goes back to <next.event>$_{Proc}$. For <hospitalisation>$_{Proc}$, the model starts by determining whether <hospitalisation>$_{E}$ was avoided (<avoided.hospitalisation>$_{E}$, which is an intermediate outcome conditional on <hospitalisation>$_{E}$ that can only happen in the EWS+DA intervention). If so, <patient>$_{Ent}$ moves to <outpatient>$_{Proc}$; if not, the model records time, costs, life years, and QALYs before determining if the <patient>$_{Ent}$ dies in hospital (<death.in.hospital>$_{E}$, which is also an intermediate outcome conditional on <hospitalisation>$_{E}$). If he does, <patient>$_{Ent}$ moves to <death>$_{Proc}$; if not, the model updates attributes and the <patient>$_{Ent}$ goes back to <next.event>$_{Proc}$. In <death>$_{Proc}$ the model follows these sequential steps: (i) recording time, costs, life years, and QALYs, (ii) updating attributes, (iii) computing total outcomes for the simulation, and (iv) removing <patient>$_{Ent}$ from the simulation (see Figure 5.1 for a diagrammatic representation of the model structure).

Each <patient>$_{Ent}$ created in the model runs through the simulation three times – one for each of the interventions under analysis.

Patient attributes and regression equations

A study by Pocock et al. (33) identified the following as significant independent predictors of mortality in HF patients: age, ejection fraction, NYHA class, serum creatinine, diabetes, not prescribed beta-blocker, systolic blood pressure, body mass index, time since diagnosis, smoking status, chronic obstructive pulmonary disease (COPD), gender, and not prescribed ACE-inhibitor or angiotensin-receptor blockers. These variables were all present in our dataset and were used in the model to predict time-to-death. We also used these variables to predict time-to-hospitalisation, as it seems reasonable to assume that the pathophysiological mechanisms leading to death in HF are the same that lead to hospitalisations. Table 5.2 contains the summary and the definitions of the parameters used in the regression equations and in the model.

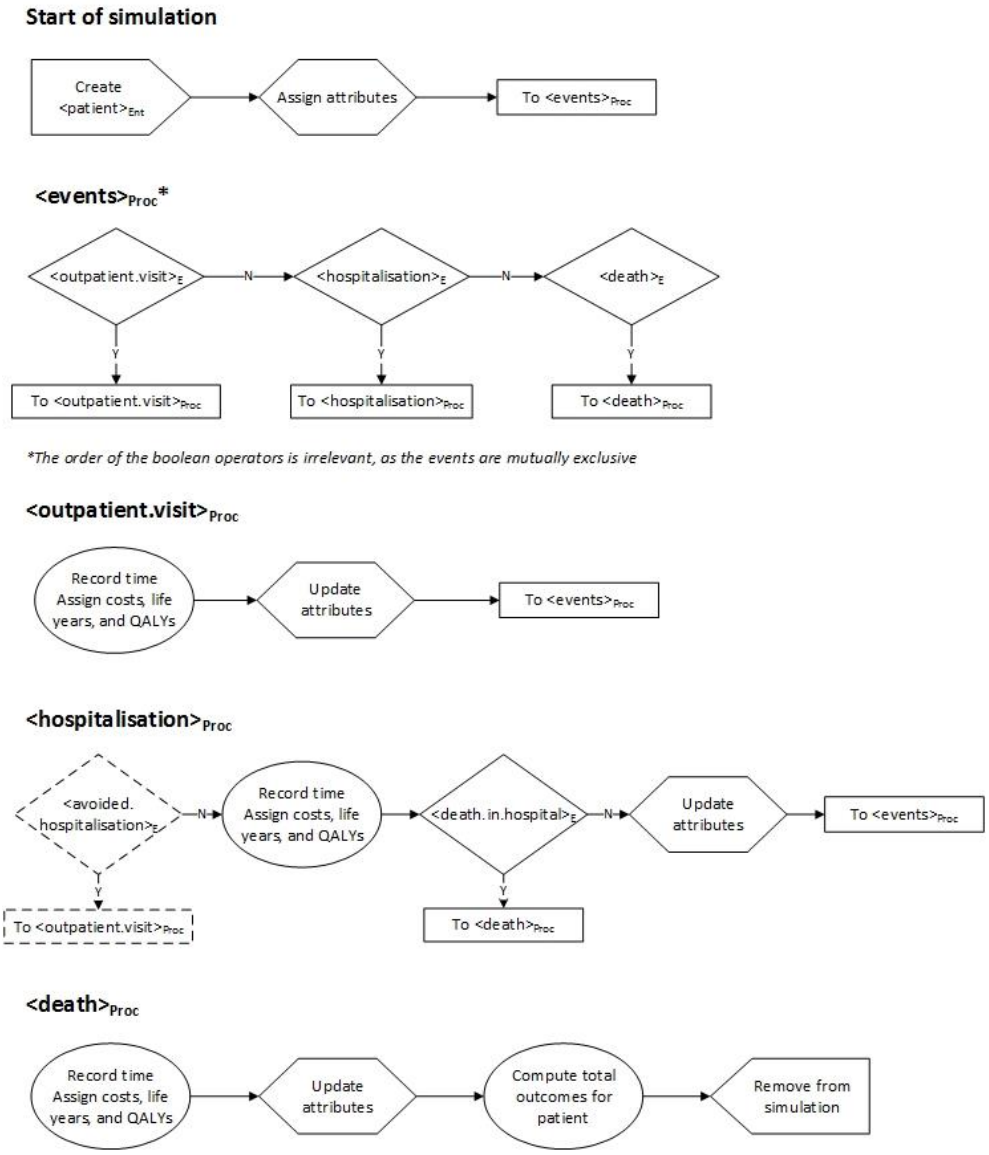


Figure 5.1 – Model structure (<avoided.hospitalisation>E, dashed in <hospitalisation>Proc, is only possible for EWS+DA)

Table 5.2 – Definition of parameters in the model

Parameter	Definition
Patient attributes	
Intervention	EWS = 1, UC = 0.
Ejection fraction (EF)	Ejection fraction (%).
Age	Age in years. Updated at every event.
Systolic blood pressure (SBP)	Systolic blood pressure in mmHg.
Body mass index (BMI)	BMI calculated as weight/height^2 (kg/m^2).
Creatinine	Serum creatinine in µmol/l.
NYHA class	New York Heart Association (NYHA) Classification I to IV (1, 2, 3, or 4).
Gender	Male = 1, Female = 0.
Smoker	Current smoker = 1, non-smoker = 0.
Diabetes	Diabetic = 1, non-diabetic = 0.
Chronic obstructive pulmonary disease (COPD)	COPD present = 1, no COPD = 0.
Recent diagnosis	Diagnosis < 18 months from baseline = 1, diagnosis > 18 months from baseline = 0.
Beta-blocker medication	Without beta-blocker medication = 1, on beta-blocker medication = 0.
Angiotensin-converting enzyme (ACE) inhibitor medication	Without ACE inhibitor medication = 1, on ACE inhibitor medication = 0.
Age x ejection fraction	Variable describing the interaction between age and the ejection fraction through the product of these variables.
Systolic blood pressure x ejection fraction	Variable describing the interaction between systolic blood pressure and the ejection fraction through the product of these variables.
Myocardial infarction	Previous history of myocardial infarction.
Chronic atrial fibrillation	Previous history of chronic atrial fibrillation.
Previous hospitalisation	Number of hospitalisations that already occurred for the simulated patient. Updated at every event.
Utility	EQ-5D-3L utility measured at baseline. Updated with utility multipliers at every event.
General model inputs (set by user)	
Number of patients	Number of patients in the simulation.
Parametric distributions	Choice of parametric distribution – exponential, Weibull, log-normal, log-logistic, and Gompertz – for time-to-death and time-to-hospitalisation calculations.

Time to outpatient visit	Time-to-outpatient visit.
Utility multipliers	Utility multipliers for updating patient utility at each outpatient visit and hospitalisation.
Discount rates	Yearly discount rates for costs and for health outcomes (life years and QALYs).
Resource costs	Yearly cost of maintenance treatment: composite costs associated with the intervention (different for UC and EWS). Alarm management costs: costs of a telephonic consultation. Event costs: individual costs for an outpatient visit, a hospitalisation, and death.
DA characteristics	
Sensitivity	Proportion of people who have the disease and are identified as having the disease, i.e. the probability of correctly detecting a hospitalisation.
False positive rate	Proportion of all the people who do not have the disease who will be identified as having the disease (= 1 - specificity).
Avoid hospitalisation	Probability of avoiding a hospitalisation in the case of having correctly predicted it.
Number of events (intermediate outcomes)	
Outpatient visits	Number of outpatient visits.
Hospitalisations	Number of effective hospitalisations.
Avoided hospitalisations	Number of avoided hospitalisation (only in the EWS+DA intervention).
Deaths	Mortality (split in hospital mortality and mortality from other causes).
Model (final) outcomes	
Costs	Total costs accrued during the simulation.
Life years	Life years accrued. Time spent in the simulation before death.
QALYs	QALYs accrued. QALYs are obtained by weighing life years with the utilities during simulation for each patient.

Abbreviations: ACE, angiotensin-converting enzyme; BMI, body mass index; COPD, chronic obstructive pulmonary disease; DA, diagnostic algorithm; EF, ejection fraction; EWS, early warning system; NYHA, New York Heart Association; QALY, quality-adjusted life year; SBP, systolic blood pressure; UC, usual care.

Time-to-event calculations

We estimated Kaplan-Meier (KM) curves for death and hospitalisation using the patient-level data for the UC and HTM populations of the TEN-HMS trial. We then fitted the most common parametric distributions – exponential, Weibull, log-normal, log-logistic, Gompertz, and generalised gamma – to the KM curves (see Appendix 5.1 for further details).

Time-to-outpatient visit (for both UC and EWS) is a model input that can be set by the user, as it may change according to the setting of the analysis, while <avoided.hospitalisation>$_E$ (see section on the conceptualisation of the DA for the details of its calculation) and <death.in.hospital>$_E$ (see section Death in hospital) are intermediate outcomes conditional on <hospitalisation>$_E$.

Death in hospital

When a patient is hospitalised, there is a chance of dying in the hospital. For predicting it, we ran a logistic regression where the probability of dying in hospital is explained by age, gender, previous history of myocardial infarction and/or chronic atrial fibrillation, comorbidities (diabetes and/or COPD), and the number of previous hospitalisations (see Appendix 5.2 for further details on the regression model).

Resource use and costs

The model distinguishes between yearly cost of maintenance treatment' for UC and for EWS, costs related to the management of false positive alarms, and event costs (outpatient visit, hospitalisation, and death). Costs of maintenance treatment and alarm management depend on the time elapsed between simulated events and are continuously discounted, while event costs are accounted for at time of occurrence and are discretely discounted.

Utilities

Utility is a patient attribute assigned at the start of the simulation according to the NYHA class at baseline. The mean utility values per NYHA class used were reported elsewhere (205) (0.88, 0.71, 0.61, and 0.49 for NYHA classes I, II, III, and IV, respectively). Every time an outpatient visit or a hospitalisation is processed, the patient utility is updated via a multiplier. For instance, if the utility at start of the simulation is 0.80 and the multiplier for hospitalisation is 0.85, the updated utility of that patient after being hospitalised is 0.80 x 0.85 = 0.68, which remains the utility for the patient until the next event is processed. The decrease in utility in the simulation is limited to the utility found for NYHA class IV.

Model outcomes

The following outcomes are calculated from the model: number of events per type (referred to as intermediate outcomes), total costs, total life years, total QALYs, and incremental cost-effectiveness ratios (ICERs).

The costs in the model are calculated through adding the discrete costs for each event (outpatient visit, hospitalisation, and death) and the cost of maintenance treatment for the intervention. Life years correspond to the elapsed time between the creation of the patient and his death and consequent removal from the simulation. QALYs are obtained through weighing life years with patient utilities over time. The ICERs were calculated as the difference in the total average costs per patient divided by the difference in the average number of QALYs per patient (€/QALY) between two alternative treatment options.

Since outcomes are recorded for each simulated patient, the model allows for extracting the individual patient history for every simulation. See Table 5.2 for a summary of the parameters used in the model.

Base-case analysis

The base-case number of simulations in the deterministic analysis was set to 1,000 patients, as this number gave stable results while keeping the running time reasonable. For the base-case analysis, the Weibull distribution was used for extrapolating time-to-death and the log-normal distribution for extrapolating time-to-hospitalisation. Distributions were chosen according to the recommendations issued by the Decision Support Unit commissioned by The National Institute for Health and Clinical Excellence (NICE) (206) (details can be found in Appendix 5.1). The time-to-outpatient visit was set to 0.234 years (≈ every 2.8 months) for UC and 0.141 years (≈ every 1.7 months) for EWS, following the data reported in the TEN-HMS study (126). The utility multipliers were set to 1 for an outpatient visit (assuming no utility changes resulting from an outpatient visit) and 0.82 for hospitalisation, which corresponds to the decrease in utility resulting from a transition from NYHA class 3 to 4 that was found in a previous study estimating QALY weights based on NYHA functional class in an elderly population with HF (207). The sensitivity of the DA was set to 0.96 and the false positive rate to 0.54, representing the Youden-point of the ROC curve provided by the manufacturer. The probability of avoiding a hospitalisation in the case of having correctly predicted it was set to 0.5, as reported elsewhere (208). A summary of input costs and respective sources is presented in Table 5.3. The costs are reported in euros and adjusted to 2020 rates based on the Dutch consumer price index (209). The costs presuppose a healthcare perspective, since it is likely that in the Netherlands there will be health care insurers that will decide upon the

availability of EWS to patients (81). Costs and health outcomes were discounted at 4.0% and 1.5%, respectively, according to Dutch guidelines (210).

Table 5.3 – Input costs in base-case analysis

Item	Estimate (€)	Source
UC outpatient visit	46.33	iMTA costing tool (211)
EWS outpatient visit	44.63	iMTA costing tool (211)
Hospitalisation	4,937.36	Stevanovic 2014 (212)
Death	1	Assumption (set to 1 for allowing PSA)
Management of false positive alarm	18	iMTA costing tool (211)
UC cost of maintenance treatment per year	705.71	Grustam 2018 (205)
EWS cost of maintenance treatment per year	2,621.70	Grustam 2018 (205)

Abbreviations: EWS, early warning system; iMTA, Institute for Medical Technology Assessment; PSA, probabilistic sensitivity analysis; UC, usual care

Probabilistic sensitivity analysis

In addition to the patient heterogeneity stemming from the variation in the patient population at baseline, the model includes two other types of uncertainty: (i) stochastic uncertainty, which is the uncertainty owing to the randomness of drawing values from probability distributions during the simulation and (ii) parameter uncertainty, which is the uncertainty associated with the coefficients of the regression equations and with the remaining model input parameters.

Accounting on the above, the probabilistic sensitivity analysis was implemented as a double loop: an inner loop in which a pre-determined number of patients are sampled with replacement from the baseline population, and an outer loop in which values of the input parameters of the model are randomly drawn. This approach is similar to other published and validated patient-level simulation models (213).

Model development, coding, and validation

The model was developed using R software (214) and it consists of four R files: (i) the survival analyses, (ii) the logistic regression model for calculating the probability of a patient dying in hospital, (iii) the model functions, which can be seen as the model engine, and (iv) the model script where the user can define the model inputs, run the

model, and output results. The full code can be found on GitHub (https://github.com/fernandoalbuquerquealmeida/EWS_HF_DES_model).

We used the Assessment of the Validation Status of Health-Economic decision models tool (AdViSHE) for having a structured view on the main topics regarding the validation of the model (215).

Results

Base-case analysis

Table 5.4 shows the average model results per patient over lifetime. UC patients experienced on average 3.61 outpatient visits per year and 1.69 hospitalisations per year, with an average cost of 17,191€ over 2.07 life years (1.19 QALYs). 43.2% of these patients died in hospital and the remaining 56.8% died from other causes. Patients treated with the EWS experienced on average 6.60 outpatient visits per year and 1.67 hospitalisations per year, with an average cost of 28,440€ over 2.88 life years (1.64 QALYs). 61.5% of them died in hospital and 38.5% from other causes. Patients who had the DA added to the EWS lived on average 3.80 years (2.21 QALYs) with an average cost of 38,120€ over that period. During that same period, patients experienced 6.58 outpatient visits per year, 1.02 hospitalisations per year, and avoided being hospitalised 0.93 times per year. 47.4% of them died in hospital and 52.6% from other causes.

Table 5.4 – Model results for the base-case analysis

Average outcomes per patient	**UC**	**EWS**	**EWS+DA**
Events (per year)			
Outpatient visits	3.61	6.60	6.58
Hospitalisations	1.69	1.67	1.02
Avoided hospitalisations	-	-	0.93
Death type			
Death in hospital, %	43.2	61.5	47.4
Death (other), %	56.8	38.5	52.6
Final outcomes			
Total costs, €	17,191	28,440	38,120
95% confidence interval*	[13,390 – 22,904]	[20,898 – 34,036]	[28,799 – 45,197]
Total life years	2.07	2.88	3.80
95% confidence interval*	[1.58 – 2.89]	[2.32 – 3.85]	[2.96 – 5.05]
Total QALYs	1.19	1.64	2.21
95% confidence interval*	[0.94 – 1.72]	[1.37 – 2.27]	[1.79 – 3.07]
*ICERs***			
EWS vs. UC, €/QALY		25,367	
EWS+DA vs. UC, €/QALY		20,522	

EWS+DA vs. EWS, €/QALY	16,794

Abbreviations: DA, diagnostic algorithm; EWS, early warning system; ICER, incremental cost-effectiveness ratio; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life year; UC, usual care.

* The 95% confidence intervals lower and upper bounds are the 5th and 95th percentiles, respectively, resulting from a PSA with an inner loop of 200 patients and an outer loop of 200 iterations.

** EWS is extendedly dominated by EWS+DA.

Model validation

The validation of the model outcomes found a slightly higher mortality for the simulated population when compared to the available data from the TEN-HMS trial (126): 52.8% and 40.8% in our simulation versus 51.0% for UC and 34.0% for EWS at day 450 in the trial. The percentage of estimated deaths in our simulation was also slightly higher than what would be predicted using the model published by Pocock et al. (33). The percentage of deaths after one year in our population estimated by the Kaplan-Meier method was 37.8% for UC and 23.8% for EWS. A population with these 1-year probabilities of death in the model estimated by Pocock et al. (33) would have a 3-year probability of death between 69.2 and 72.5% for UC and 49.0 and 52.3% for EWS. The estimated probabilities of death after three years in our simulation were 77.5% and 65.4%, respectively. In spite of this observation, it should be stressed that comparing mortality with the figures published by Pocock et al. (33) should not yield exactly the same results, as the considered populations are not exactly the same, both in terms of the patient characteristics at baseline, which are predictors of their survival, and the sample size generating the results. It is still worthwhile mentioning that the direction of the impact of the predictors for mortality in our model was the same as observed by Pocock et al. (33) for all variables except smoking and time of diagnostic. In our model, smoking was associated with a lower probability of dying (although with almost no effect), as well as the time since first diagnosis of HF time being lower than 18 months (see Appendix 5.1 for further details).

There were 1.69 hospitalisations per life year in the UC population and 1.67 hospitalisations per life year in the EWS population observed in the model. These hospitalisation rates were about one third higher than those observed in the TEN-HMS trial (126) (1.25 and 1.22, respectively for UC and EWS). The increased hospitalisation rates can be partly explained by the additional survival considered in the model when compared to the TEN-HMS trial, especially when weighing in the fact that increased age reduces time-to-hospitalisation, and by the lower time-to-outpatient visit used in the base-case analysis when compared to the input used for selecting the parametric model (see Appendix 5.1 for further details).

When comparing the outcomes of the model to other models addressing similar problems, we found comparable deterministic results to the ones found by Grustam et al. (205). It should be noted, however, that their study did not estimate the (cost)-effectiveness of EWS+DA. The comparisons between total costs, life years, and QALYs for UC and EWS are shown in Table 5.5.

Table 5.5 – Outcome comparison with Grustam et al. (28)

Outcome	**Present study**	**Grustam et al. (205)**	**% difference**
Total costs EWS	28,440 €	27,186 €	4.61%
Total costs UC	17,191 €	14,414 €	19.27%
Total LYs EWS	2.88	4.02	-28.36%
Total LYs UC	2.07	2.71	-23.62%
Total QALYs EWS	1.63	2.93	-44.37%
Total QALYs UC	1.19	1.91	-37.70%

Abbreviations: EWS, early warning system; QALY, quality-adjusted life year; LY, life year; UC, usual care.

For a systematic overview on the topics related to the model validation, please consult the filled in AdViSHE questionnaire (Online Appendix 5.1).

Discussion

This study aimed at developing a health economic patient-level simulation model for HF that included a wide variety of HF patient characteristics and that simulated changes in these characteristics and their subsequent impact on a broad set of outcomes. The modelling framework should be able to model patients managed with an early warning system, with or without the use of a diagnostic algorithm.

We had access to a comprehensive patient-level dataset generated in the TEN-HMS study (126) that contained the critical factors for prognosis as identified previously by Pocock et al. (33). The limitations of the database consisted of the relatively small sample size, the inevitable missing data on some of the variables, and referring to 2005 (126), which can overlook the changes in clinical practice that occurred ever since (216). However, it ought to be mentioned that patient-level simulation modelling in R has the clear advantage of allowing the adaptation of the code for using other available databases – as long as they include the patient and disease characteristics used in the model – for estimating the regression equations and/or for performing an external validation of the model results without changing the core model structure.

In total, we included 20 patient and disease characteristics and 8 different outcomes in the model, which allowed for an adequate description of HF patients across their treatment pathway until death. These characteristics make our model unique, as to the best of our knowledge there are not any previously published models in HF that are able to take into account individual patient characteristics for generating suitable outcomes for our target population (59). Disease pathways and health outcomes in HF – alike other chronic diseases – are strongly influenced by the individual characteristics of the patients (12, 33, 217-219). It is therefore crucial that the type of model chosen allows for recording the individual patient experience and the variation of their individual characteristics over time. In this regard, Markov models have three critical shortcomings when compared to patient-level simulations: (i) the definition of health states may preclude considering inter-patient variability, (ii) the fixed cycle length does not allow for exploring the effects of changing the frequency of events that impact individual patient characteristics (e.g. outpatient visits), and (iii) the "lack of memory" regarding the treatment history of a patient when in fact the treatment options of chronic patients normally depend on the previous treatment sequencing and experiences with those treatments (220). Conversely, DES models can address a wide range of problems, as health economic modelling using events is a more flexible approach than using health states. Further, DES models use patient attributes, which can change over time and affect time-to-event calculation, to properly model competing risks. Since the DES models approach patients individually, they are a

better alternative for dealing with heterogeneous populations. DES models are perceived as a better option for conveying the message to non-modelling experts, as consist of a more compact representation of the conceptual model, avoiding, for instance, the problem of overcomplicated Markov chains through state explosion. Further, in the eventuality of limited data, DES models also provide a substantial advantage, as the inadequacy of the data is not built into the structure of the model; the simulation can be designed to properly reflect the problem under analysis and carry out exploratory analyses with limited data and best-guess estimates (221-224). So, although there is a need of a detailed and comprehensive database for estimating the regression equations governing the time-to-event calculations, after the development and validation of the model, which was the goal of our study, it is possible to test a wide variety of scenario and perform subgroup analyses by changing the settings of the model and/or the simulated model population.

Building on the specific features of DES modelling, it is of the utmost importance to stress the ability of our model to estimate health outcomes for the EWS+DA intervention, with particular attention to its DA feature. In an EWS setting, clinical information is usually assessed by a clinical team who is prompted to act based on clinical decision rules defined for specific combinations of the monitored parameters as well as the assessment of the clinical picture at any given time. However, evidence shows that data-driven approaches like DAs looking at trends and patterns of recorded parameters change seems to improve the accuracy of detecting events when compared to clinical decision rules (171, 173, 225, 226). When taking into account the conceptualisation of the DA (see Methods section), since the model only needs a figure for sensitivity and specificity for accounting for the DA, it easily allows for analysing the (cost)-effectiveness of the EWS+DA intervention at any given point the receiver operating characteristic (ROC) curve of the DA. In other words, the model permits judging on the best operating point for the DA in order to optimise the cost-effectiveness of the intervention, which is crucial for making informed decisions on the adoption of a particular DA. Additionally, we can think of our model as a bridge between cost-effectiveness and the huge potentialities of artificial intelligence and machine learning for improving the quality of those decisions. Not only by reducing uncertainty through the continuous incorporation of big data collected by the EWS and/or other data sources, but also by constantly improving the DA prediction capabilities through machine learning, thereby determining the best follow-up actions from the results of the DA (70, 71). We can further envision a more comprehensive model to which our model is only but a piece that is generating the cost-effectiveness results. Going one step deeper, we can think of the cost-effectiveness results themselves as another piece of information used by the DA for improving its predictions.

Although it reflects the disease pathways in HF and uses home telemonitoring (HTM) as an example of an EWS, the model was developed in order to be easily adaptable for other type of EWS interventions used in chronic disease management. For instance, the time-to-outpatient visit, which can be easily changed in the model by the user, can be set according to the specific treatment guidelines for any given population suffering from a chronic disease. In our case, the EWS had an effect in both time-to-hospitalisation and time-to-death. However, other events can be considered when conceptualising the model for other chronic diseases; the logic used for modelling hospitalisation and death in our model can be repeated for as many events as needed. Focusing on the DA, it should be noted that this feature affected the outcomes of the simulated patient by avoiding hospitalisations (having an impact in costs and health outcomes). Avoiding hospitalisations, in turn, affects the disease pathways of the simulated patient and has an impact on recorded outcomes. This logic can be used with other EWS for events a DA is intended to avoid in the management of any other chronic disease.

Concerning the validation of the model, the face validity of the conceptual model was underpinned by the opinions of both experts in the field of health economic modelling and a multidisciplinary team of experts in the field of clinical technical solutions development for HF. All the performed tests revealed that our model was robust and able to generate health outcomes comparable to those estimated by other models addressing similar problems and those obtained from empirical data. On the comparison to other models, it should be stressed that we found fewer life years and QALYs than Grustam et al. (205). In their study, the authors assumed that the transition probabilities measured in the time frame of 240 to 450 days in the original study continue unaltered for 20 years. Given the average age of the patients included in the model (67 years old) and their very poor health state, it seems unlikely that their transition probabilities would remain the same for the following 20 years. Therefore, the fewer QALYs found in our study are a consequence of the higher mortality that was found using the parametric survival modelling approach that we took and the assumption that there is a utility change similar to the one observed for a change to the next worse NYHA class from a hospitalisation, which occurs more frequently than the health state transitions in the study by Grustam et al. (205). The AdViSHE questionnaire proved to be a useful tool in the process of model validation, both for guiding in the model development and for identifying areas for improvement (see Online Appendix 5.1). On that note, there are a few shortcomings of the model that ought to be discussed.

Although the model allows for updating patient characteristics at the occurrence of each event, we did not have information on the evolution of some patient characteristics and we could not update patient attributes accordingly. Conceptually it

would be ideal to have equations describing the trends of the patient characteristics, eventually with a link to changes in the medication that could be modelled during the outpatient visit procedure.

The model outcomes are representative for the group of patients who participated in the TEN-HMS trial, which are mainly severe HF patients that have been previously hospitalised. It ought to be said that HF patients enrolled in clinical trials of EWS usually have similar characteristics to the TEN-HMS patients and, as such, results could be projected for those patients using the model. However, since regression equations were estimated using the database obtained from the TEN-HMS trial, extrapolation of the results to the general HF population should be done with care. It would be interesting to re-estimate the model equations using real world evidence for a more representative HF population in order to assess whether there are significant differences in estimated outcomes. In doing so, the model would be able to be used for a larger proportion of HF patients – for example, a HF population with milder symptoms and treated in primary care – who could also be candidates for an EWS. However, it should be noted that building a DES model is an extensively data-demanding exercise that requires a wide range of patient-level data for building and validating the model. Unfortunately, patient-level data are not widely available, particularly in the real world setting, and they tend to be characterised by a lot of missing data, which leave the developer with a dilemma on how to handle those without biasing the outcomes of the model (227-231).

Further on the issue of data, in our particular case, we did not have information that would allow us to determine the impact of patient characteristics in outpatient visits. If we would have been able to do so, we could have incorporated in the model a relationship between patient characteristics and outpatient visits, which could result, for instance, in a change in medication. The change in medication in turn could impact the disease pathways in the model and, as a consequence, the outcomes of simulated patients. This would arguably be of added value from a conceptual point of view and for the sake of increased face validity of the model in the eyes of the layperson in health economics – as it is often the case of some decision-makers.

We also regret not having access to another database with patient-level data, which would have been worthwhile for increasing the sample size of our data inputs (thus reducing uncertainty) and for validating the model through assessing outcomes using alternative input data. Yet again, data availability and the real world hardly go hand-in-hand.

In conclusion, the developed model is a unique patient-level simulation model that includes many of the patient and disease characteristics that are considered important for prognosis and/or treatment of HF patients. The model can be used for simulating a

wide range of outcomes for different patient subgroups. More specifically, the model can provide useful information for guiding research and for the development of new treatment options, with a particular focus on early warning systems and the operationalisation of diagnostic algorithms, by showing the possible impact of these interventions on a large number of important HF outcomes.

Appendices

Appendix 5.1 – Survival analyses and time-to-event calculations

Appendix 5.1 presents the results concerning the survival analyses for death and hospitalisation. The detailed explanation of the performed analyses is only presented for time-to-death (Appendix 5.1-A). The interpretation of the time-to-hospitalisation analyses (Appendix 5.1-B) is briefly described and it should to be complemented by the explanations presented in the time-to-death analyses considering hospitalisation as the survival event of interest instead of death.

Appendices 5.1-A and 5.1-B present information on the following analyses: (1) Kaplan-Meier estimates, (2) survival curves difference (log-rank or Mantel-Haenszel test and Peto & Peto modification of the Gehan-Wilcoxon test), (3) Cox proportional hazards model, (4) parametric survival modelling, and (5) time-to-event estimates derivation from parametric survival models.

Appendix 5.1-A – Time-to-death

1. Kaplan-Meier estimates

The Kaplan–Meier (KM) estimate is a non-parametric statistic used to estimate the survival function through measuring the fraction of patients who have not experienced an event of interest at each point in time since the start of follow-up.

Suppose that events occur at D distinct times $t_1 < t_2 < \cdots t_D$, at time t_i there are d_i events (e.g. deaths), and r_i is the number of individuals who are at risk (of death) at time t_i. Then, the quantity d_i/r_i is an estimate of the probability (p_i) that an individual who survives to just before time t_i experiences the event at that time.

The complement of p_i ($\widehat{p_i} = 1 - d_i/r_i$) is the estimated probability of surviving day i given survival until day i. The survival probability at time D is thus calculated through the formula $\hat{S}(D) = \widehat{p_1} \times \widehat{p_2} \times \ldots \times \widehat{p_D}$. Plotting the survival probabilities calculated at all data points for which there is an observed event or censoring (i.e. observation that ended before occurrence of the event) results in the survival function.

The Kaplan-Meier method makes no (parametric) assumptions on the shape of the underlying survival or hazard curves. The resulting graph resembles a two-dimensional downward staircase starting at survival probability 1 (when all patients are alive) and dropping over time according to the events observed during follow-up. Censoring is represented in the KM graph by a vertical dash on the curve representing the survival function. The KM estimate for death for the usual care (UC) and the early warning system (EWS) interventions is presented in Figure 5.2.

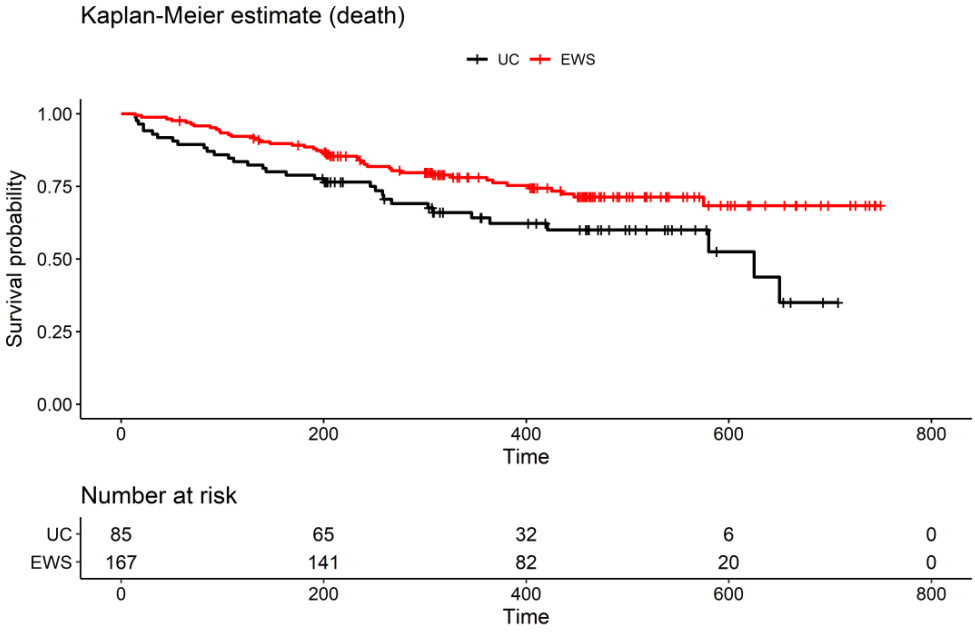


Figure 5.2 – Kaplan-Meier estimates for death (time in days)

2. Survival curves difference

The purpose of the survival curves difference analyses is to test whether two survival curves are statistically different, i.e. if one intervention is better than the other in terms of survival probability. Thus, considering that the "survival" event is death, in our example, EWS is better than UC if the survival time with EWS is larger than survival time with UC. For carrying out this test, we define the null hypothesis as no difference in survival probabilities: $H_0: S_{EWS}(t) = S_{UC}(t)$ and we run global tests comparing the two curves over the whole time range.

We used both the log-rank or Mantel-Haenszel test and the Peto & Peto modification of the Gehan-Wilcoxon test.

a. log-rank or Mantel-Haenszel test

The log-rank or Mantel-Haenszel test is based on the comparison of observed number of events (O) in index group with the expected number of events (E) if the null hypothesis were true. For each event time point, a 2 x 2 table is made and the expected no of events in the index group is determined (see Table 5.6).

Table 5.6 – 2x2 table for number at risk and number of events for UC and EWS

Group	Number at risk	Number of events
EWS	r_{EWS}	d_{EWS}
UC	r_{UC}	d_{UC}
Total	$r_{EWS} + r_{UC}$	$d_{EWS} + d_{UC}$

Expected events in the EWS group at event time point t are given by $E_{EWS} = \frac{r_{EWS}(d_{EWS}+d_{UC})}{r_{EWS}+r_{UC}}$. The analogue is also true for the UC group.

E is the sum of all these expected numbers over all time points (2x2 tables). The log-rank test statistic is obtained by summing the χ^2 statistics of the 2x2 tables over all event time-points and it is given by $Z = \frac{O_{tot}-E_{tot}}{sd_{tot}}$.

Under H_0, Z follows a standard normal distribution. Often Z^2 is given, which follows a χ^2 distribution with 1 degree of freedom under H_0. Since we are dealing with differences, it does not matter which is the reference group and which is the index group, as Z^2 will remain the same.

The actual values of the time-points are irrelevant – only their ordering (rank test) –, as the log-rank test is formally identical to the Mantel-Haenszel test for a series of 2x2 tables. The log-rank test has optimal power in case the ratio of the hazards is constant over time, but it has little or no power if hazards cross. Table 5.7 presents the results of the test.

Table 5.7 – log-rank or Mantel-Haenszel test for death

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
UC	85	33	23	4.31	6.24
EWS	167	42	52	1.91	6.24
Chisq = 6.2 on 1 degrees of freedom, p = 0.01					

The results show a statistically significant difference between the survival curve of UC and EWS, i.e. a difference in survival for both interventions considered.

b. Peto & Peto modification of the Gehan-Wilcoxon test

The Gehan-Wilcoxon test is a variation of the log-rank test statistic and it is derived by applying different weights at consecutive failure times. Peto-Peto modifications are characterised by nonparametric generalised maximum-likelihood estimates of the survival function for interval-censored data; they are useful in early differences and more robust in situations where many observations are censored.

The combined approach of these methods is most sensitive to early differences (or earlier time points) between survival, while the previously described log-rank is most powerful under proportional hazards. The Peto & Peto modification of the Gehan-Wilcoxon test is therefore also used to assess whether the effect of the treatment on survival is strongest in the earlier phases of administration and if it tends to be less effective over time. Table 5.8 shows the results of the test.

Table 5.8 – Peto & Peto modification of the Gehan-Wilcoxon test for death

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
UC	85	28.1	19.6	3.71	6.28
EWS	167	35.2	43.7	1.66	6.28
Chisq = 6.3 on 1 degrees of freedom, p = 0.01					

The results reveal a statistically significant difference between the survival curve of UC and EWS, i.e. a difference in survival between the considered interventions.

3. Cox proportional hazards model

We are dealing with survival data and we have previously showed a non-parametric way of estimating the survival curve. We have also discussed the problem of testing whether two (or more) survival curves are equal (log-rank test and the Peto & Peto modification of the Gehan-Wilcoxon variation) and thus inferring on the effect of any given intervention. Now we will focus on the quantification of the effect of covariates on survival.

In other words, we would like to have an effect size of the intervention, not only a P-value, when comparing two survival curves. Additionally, we would like to study the effect of continuous covariates (e.g. age) and dichotomous covariates (e.g. gender) on survival. We are interested at looking at several covariates at the same time.

The postulated problem is linked with regression analysis in its general sense, i.e. the set of statistical processes used for estimating the relationship between a dependent variable (outcome variable) and one or more independent variables (covariates or explanatory variables). The statistical relationship between the covariates and the outcome variable is studied though regression models. The type of regression model depends on the distribution of the outcome variable given the covariates.

In the particular case of an outcome variable being survival data, we frequently use a regression model referred to as a Cox proportional hazards model for quantifying the impact of the covariates on survival. We will now formalise it mathematically:

- In the Cox proportional hazards model, if $h_1(t)$ is the hazard rate of treatment A and $h_2(t)$ is the hazard rate of treatment B, then the ratio of these hazards – the hazard ratio – is $HR(t) = \frac{h_2(t)}{h_1(t)}$.
- Since $h_1(t)$ and $h_2(t)$ both depend on time, $HR(t)$ also depends on time. The proportional hazards assumption that gives name to the model states that $HR(t)$ does not depend on time and it is a constant.
- If Z is the covariate referring to the intervention and we define $Z = 0$ as the usual care (UC) and $Z = 1$ as the early warning system (EWS), then $h_0(t)$ is the hazard rate corresponding to the reference category (UC) – also called baseline hazard. Then the model is defined by $h(t|Z) = h_0(t)\exp(\beta Z)$, where:
    - $Z = 0$: $h(t|Z = 0) = h_0(t)\exp(\beta 0) = h_0(t)$ – hazard rate of UC (Z = 0); above called $h_1(t)$;
    - $Z = 1$: $h(t|Z = 1) = h_0(t)\exp(\beta 1) = h_0(t)\exp(\beta)$ – hazard rate of EWS (Z = 1); above called $h_2(t)$;
    - and the hazard ratio is given by $\frac{h(t|Z=1)}{h(t|Z=0)} = \frac{h_0(t)\exp(\beta)}{h_0(t)} = exp(\beta)$.
- The relationship between the covariate and the hazard is given by the expression:

$h(t|Z) = h_0(t)\exp(\beta Z)$.

- The relationship between the covariate and the hazard is given by the expression:

$H(t|Z) = H_0(t)\exp(\beta Z)$.

- And the relationship between the covariate and the survival function is given by the expression:

$S(t|Z) = \exp\big(-H(t|Z)\big) = \exp(-H_0(t)\exp(\beta Z)) = S_0(t)^{\exp(\beta Z)}$.

a. Parameter estimation

We fist estimated the hazard ratio between the interventions by fitting a Cox proportional hazards model where the intervention was the only covariate (see results in Table 5.9).

Table 5.9 – Cox proportional hazards regression model results for death (intervention only)

covariate	exp(coefficient)	Standard error	P-value
intervention	0.563	0.233	0.014
n = 252, number of events = 75 Likelihood ratio test = 5.86 on 1 df, p = 0.015			

The results of this model show a statistically significant hazard ratio between the EWS and UC of 0.563. In other words, a patient treated with the EWS has a 43.7% decrease on the hazard of dying when compared to the UC intervention.

We subsequently estimated the impact of multiple covariates in survival through another regression model. A definition of the covariates used in the model is presented in Table 5.10.

Table 5.10 – Covariates used in survival regression models for both death and hospitalisation

Covariate	**Variable type**	**Definition**
intervention	Categorical	EWS = 1; UC = 0
ejection.fraction	Continuous	Ejection fraction (%)
age	Continuous	Age (years)
sbp	Continuous	Systolic blood pressure (mmHg)
bmi	Continuous	Body mass index calculated as weight/height^2 (kg/m^2)
creatinine	Continuous	Serum creatinine (µmol/l)
(nyha.class)2	Categorical	NYHA class 2 (Yes= 1; No = 0)
(nyha.class)3	Categorical	NYHA class 3 (Yes= 1; No = 0)
(nyha.class)4	Categorical	NYHA class 4 (Yes= 1; No = 0)
gender	Categorical	Male = 1, Female = 0
smoker	Categorical	Smoker = 1; Non-smoker = 0
diabetes	Categorical	Diabetic = 1; Non-diabetic = 0
copd	Categorical	COPD present = 1; No COPD = 0
recent.diagnosis	Categorical	Diagnosis < 18 months = 1; Diagnosis > 18 months = 0
no.beta.blocker	Categorical	Without beta-blocker medication = 1; On beta-blocker medication = 0
no.ace	Categorical	Without ACE inhibitor medication = 1; On ACE inhibitor medication = 0
age.ef	Continuous	Interaction between age and ejection fraction (product of variables above)
sbp.ef	Continuous	Interaction between systolic blood pressure and ejection fraction (product of variables above)

Categorical variables in R are referred to as factors. In the results of the model shown in the R console, the covariate intervention is labelled as *factor(intervention)1*, meaning that the results for the coefficients are given for a value of 1 for the covariate intervention.

For the purpose of running regression models in R, continuous variables should be normalised. Normalisation in R is referred to as scaling. In the results of the model shown in the R console, the covariate age is be labelled as *scale(age)*, meaning that the results for the coefficients are referring to the normalised age variable.

For the purpose of simplifying the results shown, *factor* and *scale* were omitted from the covariates.

The Cox proportional hazards model with all covariates is thus defined by:

$$h\left(t\middle|Z_{intervention}, \dots, Z_{sbp.ef}\right) = h_0(t)\exp\left(\beta_{intervention}Z_{intervention} + \cdots + \beta_{sbp.ef}Z_{sbp.ef}\right).$$

The mathematical notation and the parameter estimation with increasing number of covariates can become overwhelming. Luckily, R does the latter calculations in the blink of an eye. The results of the regression model for death with all covariates are shown in Table 5.11.

Table 5.11 – Cox proportional hazards regression model results for death (all covariates)

Covariate	exp(coefficient)	Standard error	P-value
intervention	0.500	0.253	0.006
ejection.fraction	0.192	1.107	0.136
age	1.415	0.612	0.571
sbp	0.354	0.394	0.008
bmi	0.625	0.168	0.005
creatinine	1.464	0.113	0.001
(nyha.class)2	1.101	0.370	0.794
(nyha.class)3	1.620	0.377	0.201
(nyha.class)4	1.605	0.507	0.351
gender	1.350	0.341	0.379
smoker	0.870	0.483	0.774
diabetes	1.330	0.272	0.295
copd	1.186	0.284	0.547
recent.diagnosis	0.748	0.268	0.279
no.beta.blocker	1.614	0.259	0.064
no.ace	1.139	0.288	0.652
age.ef	1.599	1.184	0.692

sbp.ef	4.621	0.835	0.067

n = 252, number of events = 75
Likelihood ratio test = 84.17 on 18 df, p = 1.579×10^{-10}

It is particularly interesting to notice the value of the *exp(coefficient)* for each of the covariates in the model, as it determines the impact the covariate has on survival. If the *exp(coefficient)* for a particular variable is lower than one, the presence of a characteristic (for covariates coded as categorical variables) or the increase of its value (in the case of continuous variables; e.g. ejection fraction) decrease the risk of dying and result in an increased chance of survival. For instance, in our model, a patient treated with EWS and with increasing ejection fraction has a higher chance of survival than that same patient treated with UC and with decreasing ejection fraction. The inverse rationale can be used when *exp(coefficient)* is higher than one, i.e. older patients and patients with diabetes have an increased risk of dying and a lower chance of survival.

b. Proportional hazards assumption test

The proportional hazards (PH) assumption can be checked using statistical tests based on the scaled Schoenfeld residuals. The function *cox.zph()* provides a solution for testing the PH assumption for each covariate included in a Cox PH regression model fit.

For each covariate, the function *cox.zph()* correlates the corresponding set of scaled Schoenfeld residuals with time, in order to test for independence between residuals and time. Additionally, it performs a global test for the model as a whole.

The PH assumption is supported by a non-significant relationship between residuals and time and it is refuted by a significant relationship. The results of the *cox.zph()* function for both our models – with the intervention as the only covariate and with the full set of covariates – are shown in Table 5.12 and Table 5.13, respectively.

Table 5.12 – Schoenfeld residuals of the Cox proportional hazards model covariates for death (intervention only)

covariate	chisq	df	P-value
factor(intervention)	0.0488	1	0.825
GLOBAL	0.0488	1	0.825

Table 5.13 – Schoenfeld residuals of the Cox proportional hazards model covariates for death (all covariates)

covariate	chisq	df	P-value
factor(intervention)	0.0284	1	0.866
scale(ejection.fraction)	0.1044	1	0.747
scale(age)	7.4865	1	0.006
scale(sbp)	0.2828	1	0.595
scale(bmi)	4.0492	1	0.044
scale(creatinine)	0.1788	1	0.672
factor(nyha.class)	0.3724	3	0.946
factor(gender)	0.3179	1	0.573
factor(smoker)	0.0069	1	0.934
factor(diabetes)	1.6573	1	0.198
factor(copd)	1.5037	1	0.220
factor(recent.diagnosis)	0.4389	1	0.508
factor(no.beta.blocker)	1.1147	1	0.291
factor(no.ace)	1.3210	1	0.250
scale(age.ef)	1.0251	1	0.311
scale(sbp.ef)	0.2819	1	0.595
GLOBAL	23.5180	18	0.171

From these results, we can observe that the test is not statistically significant for most of the covariates and that the global test is also not statistically significant (Table 5.13). A lack of statistical significance is also observed in the test for the model with the intervention being the only covariate (Table 5.12). We can therefore assume that the proportional hazards assumption holds.

c. Influential observations test

To test for influential observations or outliers, we can visualize the dfbeta values using the function *ggcoxdiagnostics()*[in *survminer* package] by specifying the argument *type = "dfbeta"*. The function plots the estimated changes in the regression coefficients upon deleting each observation in turn. The results for our models are show in Figure 5.3 and Figure 5.4.

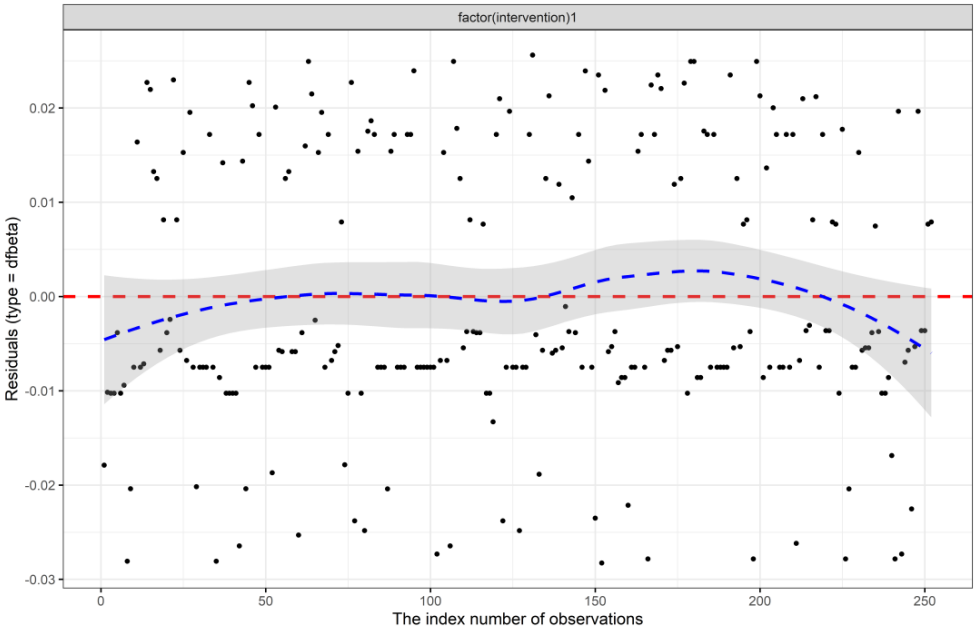


Figure 5.3 – dfbeta residuals for the Cox proportional hazards model for death with intervention as the only covariate

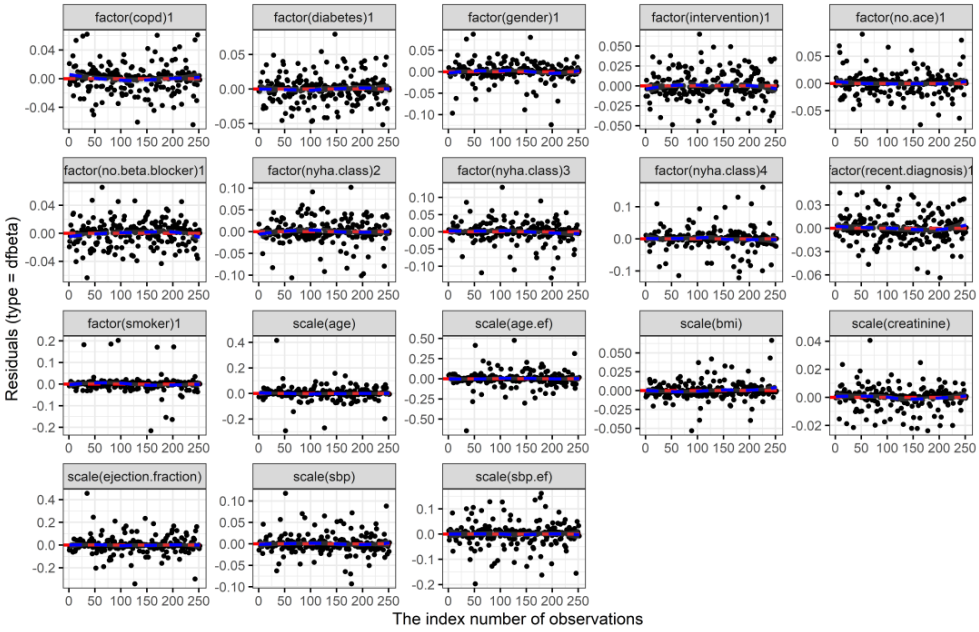


Figure 5.4 – dfbeta residuals for the Cox proportional hazards models for death (all covariates)

The index plots in Figure 5.4 show that comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that most of the observations are not much influential individually, even though some outliers (larger dfbeta values) can be spotted for some covariates. For the model with intervention as the only covariate, despite the scale of the graph changing our visual perception, the plot shown in Figure 5.3 also suggests that most of the observations are not much influential individually.

d. Log cumulative hazards

In the following section we will describe the use of parametric survival modelling for extrapolating data beyond trial duration. The consideration of the observed hazard rates over time is vital for assessing suitability of different parametric models, as they incorporate different hazard functions. For instance, exponential models are only suitable if the observed hazard is approximately constant and non-zero, while Weibull and Gompertz models allow for the incorporation of monotonic hazards, and log-logistic and log-normal models can incorporate non-monotonic hazards.

Log-cumulative hazard plots are not only used to illustrate the hazards observed in the clinical trial, but they also allow for assessing whether the proportional hazards assumption is reasonable.

Figure 5.5 shows an illustration of a log cumulative hazard plot for the Kaplan Meier curves previously shown in Figure 5.2. The graph demonstrates that there are no seemingly changes in the hazard during the trial duration and that hazards are reasonably proportional between the two treatment groups. This indicates that a single parametric model may be suitable to model survival, including a model using the exponential distribution.

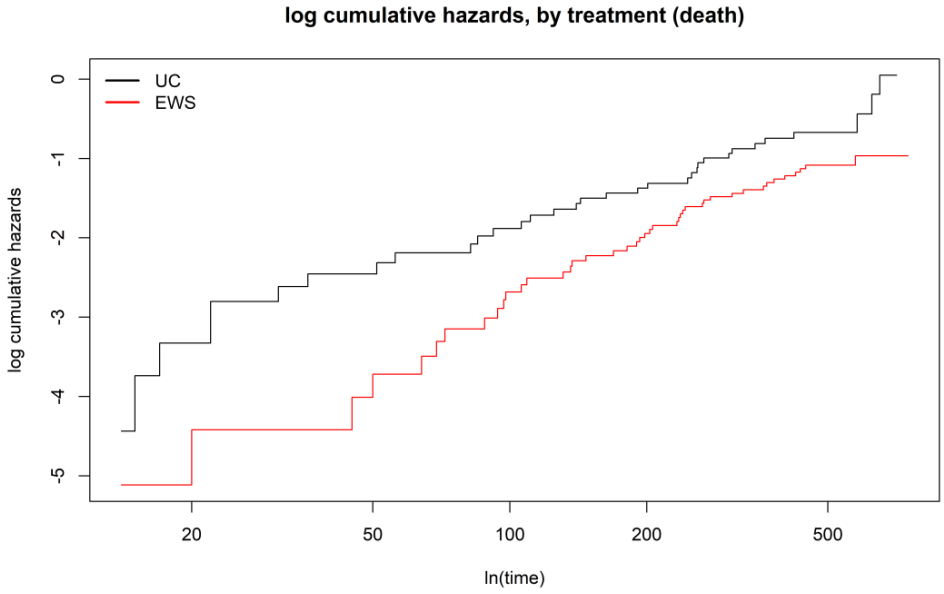


Figure 5.5 – log cumulative hazards, by treatment (death)

4. Parametric survival modelling

Survival data is different from other types of continuous data in the sense that the endpoint of interest is often not observed in all subjects, as patients may be lost to follow-up or the event may not have occurred by the end of study follow-up. Given the extreme unlikeliness of having access to complete survival data from a trial – i.e. all included patients have experienced the event by the end of follow-up –, considering that decision-making supported by health economic evaluations often requires that health effects (and costs) of medical interventions is considered over lifetime, the extrapolation of survival data is required in order to usefully incorporate them in health economic models.

Generally speaking, extrapolating survival data is achieved through the use of parametric models fitted to empirical time-to-event data. There is a wide range of parametric models available, each with its own characteristics that makes it suitable for different data sets. In our analysis, we fitted to our data the most commonly used parametric models in the context of health economic modelling: exponential, Weibull, Gompertz, log-logistic, log-normal, and generalised gamma. We modelled the intervention as a covariate since we have assumed that the proportional hazards assumptions holds based on the tests showed above. The graphical representation of the fitting of the parametric models to time-to-death data is shown in Figure 5.6.

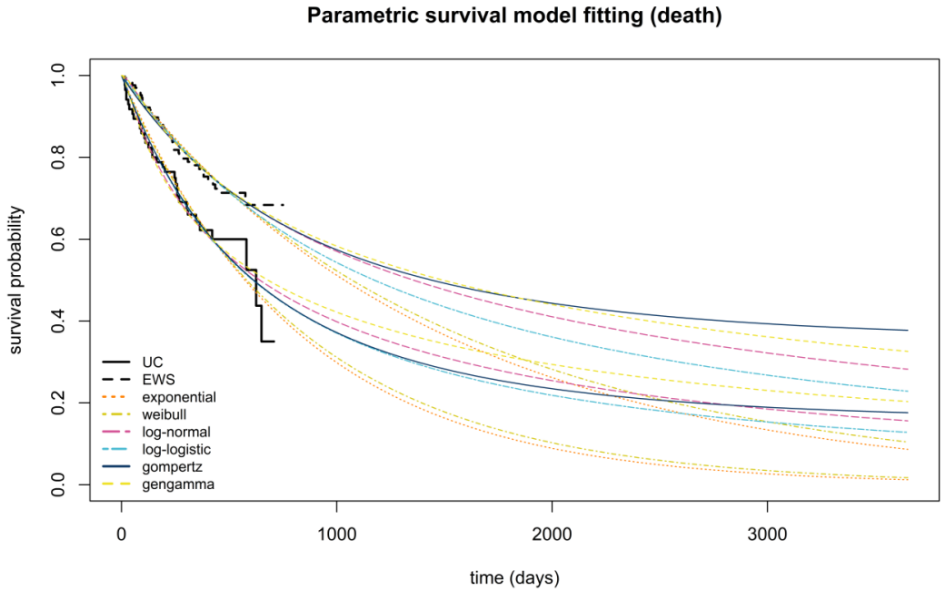


Figure 5.6 – Fitting of parametric survival models to time-to-death

Despite being the preferred method for incorporating survival data into health economic models, parametric models raise an issue on the validity of the extrapolated data. The main question becomes one of how to best make inferences about the tails of probability distributions given partial – or even completely absent – information. Special care must be taken in the common case where lifetime data are immature and non-censored observed values are only available on a small proportion of patients. As it is noticeable in Figure 5.6, extrapolations beyond trial duration using each of the considered parametric models can result in significantly different estimates for survival. These differences also get larger with increasing time.

The choice of parametric model is one of the most discussed issues in health economic modelling, as it can have a large impact on cost-effectiveness estimates and on the decisions that could stem from these estimates. Suitability assessment of each of the fitted models can be done through a wide variety of methods, which are aimed at demonstrating whether the model provides a good fit to the observed data and if the extrapolated portion is clinically and biologically plausible. We will discuss the choice of the parametric model used in the base-case analysis subsequently.

a. Visual inspection

The first step for judging how well a parametric survival model fits the clinical trial data consists in the visual assessment on how closely the graphical representation of the model follows the Kaplan Meier curves. Although this provides a simple method for choosing one model over another, it can be inaccurate and prone to unexplained variability. Further, a fitted model that follows the Kaplan-Meier curves closely may have an unlikely tail. This issue will be discussed in further detail later.

In our case there are not many striking differences between the fit of parametric model curves to the Kaplan-Meier curves (see Figure 5.6). As such, no clear-cut decision on the better suitability of a particular parametric model should be made based on visual inspection alone. We must supplement or model choice with additional testing.

b. Goodness-of-fit to the observed data

The Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are penalised-likelihood information criteria estimators that provide useful information on the relative fit of alternative parametric models for a given dataset. The AIC and the BIC consist of a goodness-of-fit term plus a penalty to control overfitting. In this way, they provide a standardised way of balancing sensitivity (having enough parameters to adequately model the relationships among variables in the population) with specificity (not overfitting a model or suggesting non-existent relationships) of parametric models.

When analysing the AIC and the BIC results, we must take into account that since some parametric models have more core parameters (i.e. excluding covariates) than others – for instance, the exponential model only has one parameter while Weibull and Gompertz have two –, multi-parameter models are more penalised in comparative terms. Also, the use of additional parameters is more highly penalised by the BIC in comparison to the AIC.

The results of the goodness-of-fit tests using the AIC and BIC for the parametric models used in our study are presented in Table 5.14.

The lay interpretation of the AIC and the BIC is the following: the lower their value, the better the fit of the parametric model to the data.

Table 5.14 – Goodness-of-fit tests for parametric survival modelling of death

Parametric model	AIC	BIC
Exponential	1211.24	1218.30
Weibull	1213.11	1223.70
Log-normal	1209.07	1219.66
Log-logistic	1211.77	1222.35
Gompertz	1212.17	1222.76
Generalised gamma	1210.82	1224.93

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion.

Judging by the small differences in AIC and BIC, one can conclude that the comparative goodness-of-fit of the alternative models to the trial data is minor. In fact, this corroborates the points made in the visual inspection of the curves, i.e. there are rather small differences between the parametric curves during the time for which there are trial data available. Again, it is worth emphasising that the AIC and the BIC only give indication about the model fit to the existing data. The extrapolation part must be taken into account using different methods.

c. Clinical validity and external data

As discussed previously, the visual inspection and the AIC/BIC tests share the crucial limitation of only being informative regarding the relative fit of the parametric models to the observed data. Despite containing useful insights, they do not provide information about the suitability of a parametric model for the time period beyond the duration of trial follow-up. In other words, the tests described in a. and b. only address the internal validity of the fitted models, but they do not allow for any consideration about the external validity of these models.

While the internal validity of a parametric model may be especially informative when survival data are fairly complete and the extrapolated portion of the parametric model should contribute little to the overall mean area under the curve, when the survival data require meaningful extrapolation it is important to validate model predictions by assessing the plausibility of the extrapolated portions of the parametric survival models. This could be done through the use of external data and/or clinical expert opinion.

External data used for judging upon the parametric model extrapolations could come from another clinical trial in a similar patient group with longer follow-up data or from long-term registry data for the relevant patient group. When these data are not available, the clinical validity of the long-term survival extrapolations could be informed by clinical expert opinion and biological plausibility. These approaches could and should be complementary.

As stated in our paper, we used patient-level data on the variables identified in the study by Pocock et al. 2013 as significant independent predictors of mortality in heart failure. This study included individual data on 39,372 patients with heart failure and it predicted 1-year and 3-year probabilities of death for different risk scores based on the independent predictors for mortality. As such, for each of those risk scores there is a correspondence between the 1-year and the 3-year probabilities of death, as shown in Figure 5.7.

Integer risk score	1-year probability of death	3-year probability of death	Integer risk score	1-year probability of death	3-year probability of death
0	0.015	0.039	26	0.175	0.397
1	0.016	0.043	27	0.191	0.427
2	0.018	0.048	28	0.209	0.458
3	0.020	0.052	29	0.227	0.490
4	0.022	0.058	30	0.248	0.523
5	0.024	0.063	31	0.269	0.556
6	0.027	0.070	32	0.292	0.590
7	0.029	0.077	33	0.316	0.625
8	0.032	0.084	34	0.342	0.658
9	0.036	0.092	35	0.369	0.692
10	0.039	0.102	36	0.398	0.725
11	0.043	0.111	37	0.427	0.756
12	0.048	0.122	38	0.458	0.787
13	0.052	0.134	39	0.490	0.815
14	0.058	0.146	40	0.523	0.842
15	0.063	0.160	41	0.557	0.866
16	0.070	0.175	42	0.591	0.889
17	0.077	0.191	43	0.625	0.908
18	0.084	0.209	44	0.659	0.926
19	0.093	0.227	45	0.692	0.941
20	0.102	0.247	46	0.725	0.953
21	0.111	0.269	47	0.757	0.964
22	0.122	0.292	48	0.787	0.973
23	0.134	0.316	49	0.816	0.980
24	0.147	0.342	50	0.842	0.985
25	0.160	0.369			

Figure 5.7 – Predicted 1-year and 3-year probabilities of death according to risk score (from: Pocock et al. 2013)

Using the Kaplan-Meier method, the 1-year probability of death in our model population was estimated to be 0.378 for the UC intervention and 0.238 for the EWS intervention. As such, we can infer on the 3-year probability of death assuming that our observed mortality corresponds to a determined risk score. Using the results reported by Pocock et al. 2013, we can match between the 1-year and 3-year probability of death and assume that the 3-year probability of death for our model population would fall between 0.692 and 0.725 for the UC and between 0.490 and 0.523 for EWS. The matching between the dying probabilities is highlighted in Figure 5.7 for UC and EWS in solid and dash, respectively.

We can then extract the predicted 3-year (1095.75 days) probability of death for each of the parametric survival models used for the UC and the EWS (see Table 5.15).

Table 5.15 – Estimated 3-year probability of death from the parametric survival models

Parametric model	UC	EWS
Exponential	0.734	0.520
Weibull	**0.721**	**0.508**
Log-normal	0.622	0.451
Log-logistic	0.652	0.482
Gompertz	0.650	0.445
Generalised gamma	0.596	0.436

From the results above, comparing to the predicted 3-year probability of death, it is fair to say that the parametric model that better estimates long-term survival (beyond available trial data) for the UC and the EWS is the Weibull (highlighted in bold in Table 5.15).

d. Conclusion

Considering the inconclusive results from the visual inspection and the goodness-of-fit to the observed data, both not allowing for making a definite choice of one model over another, we relied on the results the clinical validity tests using external data for making the Weibull our choice of the parametric model for the base-case analysis.

It should also be said that we did not have data access to long-term registry data for the relevant patient group beyond three years. From the results presented in Figure 5.6, it is perceptible that survival beyond three years can change according to the model chosen. Clinical expert opinion could have been used for informing longer-term survival, even though this is extremely fragile evidence that should be used with extreme care. Therefore, in order to analyse the impact of the choice of different parametric models on the results of the cost-effectiveness model, sensitivity analyses should be used.

5. From parametric survival models to time-to-event estimates

a. Functions used in survival analysis and how to use them in R

flexsurv is an available R package for fully-parametric modelling of survival data. The package main model-fitting function, *flexsurvreg* uses the familiar syntax of *survreg* from the standard survival package.

In order to understand the use of R and its survival packages in our model, we must start by examining the most frequently used range of parametric survival distributions, their specifications in R, and the hazard shapes they support. We will then show how the *flexsurv* package is used for the parametric regression modelling of survival data.

The primary quantity of interest in survival analysis is the survival function, defined as the probability of survival beyond time t, $S(t) = \Pr(T > t) = 1 - F(t)$, where T is a random variable denoting the time at which the event occurs. The survival function is the complement of the cumulative distribution function (CDF), $F(t) = \int_0^t f(u)du$, where $f(t)$ is the probability density function (PDF).

The hazard function, or the instantaneous rate at which an event occurs at time t given survival until time t is given by $h(t) = \frac{f(t)}{S(t)}$.

The survival function can also be expressed in terms of the cumulative hazard function, $\mathrm{H}(t) = \int_0^t h(u)du$, $S(t) = e^{-\mathrm{H}(t)}$.

R functions for the parametric distributions commonly used in survival analysis are shown in Table 5.16. The default *stats* package contains functions for the PDF, CDF, and random number generation for many of the distributions, and additional functions for the distributions that are not covered by the *stats* package, as well as the hazard functions for all distributions, are provided by the *flexsurv* package.

Table 5.16 – R functions for the parametric distributions used in survival analysis (232)

	PDF	CDF	Hazard	Random sampling
Exponential	stats::dexp	stats::pexp	flexsurv::hexp	flexsurv::rexp
Weibull (AFT)	stats::dweibull	stats::pweibull	flexsurv::hweibull	stats::rweibull
Weibull (PH)	flexsurv::dweibullPH	flexsurv::pweibullPH	flexsurv::hweibullPH	flexsurv::rweibullPH
Gompertz	flexsurv::dgompertz	flexsurv::pgompertz	flexsurv::hgompertz	flexsurv::rgompertz
Gamma	stats::dgamma	stats::pgamma	flexsurv::hgamma	stats::rgamma
Log-normal	stats::dlnorm	stats::plnorm	flexsurv::hlnorm	stats::rlnorm
Log-logistic	flexsurv::dllogis	flexsurv::pllogis	flexsurv::hllogis	flexsurv::rllogis
Generalised gamma	flexsurv::dgengamma	flexsurv::pgengamma	flexsurv::hgengamma	flexsurv::rgengamma

b. Parametrisation of distributions used in survival analysis in R

The parameterisation of the above distributions in R is shown in Table 5.17. The parameter of primary interest (in *flexsurv*; by default, also the only parameter that is dependent on the covariates) is underlined; this is known as the location parameter and it typically determines the mean or location for each of the distributions. The other parameters are ancillary parameters that determine the shape, variance, or higher moments of the distribution. The ancillary parameters impact the hazard function, which can take a variety of shapes depending on the distribution:

the exponential distribution only supports a constant hazard;

- the Weibull, Gompertz, and gamma distributions allow for monotonically increasing and decreasing hazards;
- the log-logistic and log-normal distributions support arc-shaped and monotonically decreasing hazards; and
- the generalised gamma distribution allows for an arc-shaped, bathtub-shaped, monotonically increasing, and monotonically decreasing hazards.

c. Parametric survival regression using *flexsurv* in R

In *flexsurv*, survival models are fit to the data using maximum likelihood. Each parameter can be modelled as a function of covariates z, $\alpha_l = g^{-1}(z^t\beta)$, where α_l is the lth parameter and $g^{-1}(\,)$ is a link function (usually log( ) if the parameter is strictly positive, and the identity function if the parameter is defined on the real line). By default, *flexsurv* only uses covariates to model the location parameter, although ancillary parameters can be supplied to *flexsurvreg()* using the *anc* argument.

In our study we used *flexsurvreg()* for the estimating regression models using the exponential, Weibull (AFT), log-normal, log-logistic, Gompertz, and generalised gamma distributions (passing on the *dist* argument "exp", "weibull", "lnorm", "llogis", "gompertz", and "gengamma", respectively), without supplying ancillary parameters onto the function.

Table 5.17 – Parametrisation of the distributions used in survival analysis (232)

	PDF	CDF	Hazard	Parameters
Exponential	$\lambda e^{-\lambda t}$	$1 - e^{-\lambda t}$	λ	rate = $\lambda > 0$
Weibull (AFT)	$\frac{a}{b}\left(\frac{t}{b}\right)^{a-1} e^{-\left(\frac{t}{b}\right)^{a}}$	$1 - e^{-\left(\frac{t}{b}\right)^{a}}$	$\frac{a}{b}\left(\frac{t}{b}\right)^{a-1}$	shape = $a > 0$ scale = $b > 0$
Weibull (PH)[a]	$amt^{a-1}e^{-mt^{a}}$	$1 - e^{-mt^{a}}$	amt^{a-1}	shape = $a > 0$ scale = $m > 0$
Gompertz	$be^{at}\exp\left[-\frac{b}{a}(e^{at} - 1)\right]$	$1 - exp\left[-\frac{b}{a}(e^{at} - 1)\right]$	be^{at}	shape = $a \in (-\infty, \infty)$ rate = $b > 0$
Gamma[b]	$\frac{b^{a}}{\Gamma(a)} t^{a-1}e^{-bt}$	$\frac{\gamma(a,bt)}{\Gamma(a)}$	$f(t)/S(t)$	shape = $a > 0$ rate = $b > 0$
Log-normal	$\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}$	$\Phi\left(\frac{\ln t - \mu}{\sigma}\right)$	$f(t)\ /S(t)$	meanlog = $\mu \in (-\infty, \infty)$ sdlog = $\sigma > 0$
Log-logistic	$\frac{\left(\frac{a}{b}\right)\left(\frac{t}{b}\right)^{a-1}}{\left(1+\left(\frac{t}{b}\right)^{a}\right)^{2}}$	$\frac{1}{\left(1+\left(\frac{t}{b}\right)^{a}\right)}$	$1 - \frac{\left(\frac{a}{b}\right)\left(\frac{t}{b}\right)^{a-1}}{\left(1+\left(\frac{t}{b}\right)^{a}\right)}$	shape = $a > 0$ scale = $b > 0$
Generalised gamma[c]	$\frac{\lvert Q\rvert (Q^{-2})^{Q^{-2}}}{\sigma t \Gamma(Q^{-2})} \exp[Q^{-2}(Qw - e^{Qw})]$	$\begin{cases} \frac{\gamma(Q^{-2},u)}{\Gamma(Q^{-2})}\ if\ Q \neq 0 \\ \Phi(\mathrm{w})\ if\ Q = 0 \end{cases}$	$f(t)/S(t)$	mu = $\mu \in (-\infty, \infty)$ sigma = $\sigma > 0$ Q = $Q \in (-\infty, \infty)$

[a] The proportional hazard (PH) model is a reparameterisation of the accelerated failure time (AFT) model with $m = b^{-a}$.

[b] $\Gamma(z) = \int_0^{\infty} x^{z-1}e^{-x}\,dx$ is the gamma function and $\gamma(s,x) = \int_0^{x} t^{s-1}e^{-t}dt$ is the lower incomplete gamma function.

[c] $w = \frac{(\log(t)-\mu)}{\sigma}; u = Q^{-2}e^{Qw}$.

The output of *flexsurvreg()* for each of the parametric models above consists of the estimates for each of the parameters of the distribution (see column "Parameters" in Table 5.17) and for the coefficients of each of the covariates included in the model (see Table 5.10). The output of *flexsurvreg()* also provides the lower and upper 95% confidence intervals and standard errors for all the aforementioned estimates. These measures of dispersion are crucial for running the probabilistic sensitivity analysis.

As an example, the results of the regression model transformed to the real line for the Weibull distribution is shown in Table 5.18. The details on how to obtain the location and the ancillary parameters from these results will be explained in the following section.

Table 5.18 – Results from *flexsurvreg()* for the parametric survival model using the Weibull distribution for death

Parameter/coefficient	Estimate	Lower 95% CI	Upper 95% CI	Standard error
shape	0.127	-0.068	0.322	0.100
scale	7.516	6.585	8.447	0.475
intervention	0.626	0.178	1.074	0.229
ejection fraction	1.441	-0.471	3.354	0.976
age	-0.303	-1.360	0.754	0.539
sbp	0.902	0.209	1.596	0.354
bmi	0.441	0.147	0.735	0.150
creatinine	-0.341	-0.534	-0.148	0.098
nyha.class.2	-0.121	-0.758	0.516	0.325
nyha.class.3	-0.460	-1.107	0.188	0.330
nyha.class.4	-0.478	-1.353	0.396	0.446
gender	-0.260	-0.850	0.330	0.301
smoker	0.130	-0.700	0.960	0.424
diabetes	-0.268	-0.734	0.198	0.238
copd	-0.119	-0.609	0.372	0.250
recent.diagnosis	0.251	-0.213	0.714	0.237
no.beta.blocker	-0.438	-0.896	0.020	0.234
no.ace	-0.105	-0.601	0.391	0.253
age.ef	-0.411	-2.452	1.630	1.041
sbp.ef	-1.332	-2.784	0.119	0.741

n = 252, Events: 75, Censored: 177
Total time at risk: 89965
Log-likelihood = -563.6952, df = 20
AIC = 1167.39

d. Using parameter and coefficient estimates for calculating the moments of the distributions

As an example, we will show how to calculate the moments of the Weibull parametric model for estimating death for any given patient.

The function coef() applied to a flexsurv object returns a numeric vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, where $\beta_1, \beta_2, \dots, \beta_n$ are the estimates for the parameters and the coefficients of the covariates in the model, i.e. the sequence of numbers shown in the column "estimate" in Table 5.18. Also, in the context of the discrete event simulation we can create a numeric vector $Z = (Z_1, Z_2, \dots, Z_n)$, where $Z_1, Z_2, \dots, Z_n$ are the values for a the patient characteristics which match the covariates included in the parametric model (continuous variables must be normalised; see notes in Table 5.10).

As previously discussed, we only used covariates to model the location parameter. Additionally, since both parameters in the Weibull distribution are bound to be positive, they are estimated on the log scale and must be transformed before they are passed on R functions for the parametric distributions.

We can thus calculate the survival function for a patient with the characteristics described by vector Z though calculating the moments of the Weibull distribution as follows:

- shape – the result of applying the natural exponential function to the shape estimate presented in the model results (see Table 5.18). Since shape is an ancillary parameter in the Weibull distribution it is not modelled using the covariates and it is unaffected by the patient characteristics.
- scale – the result of applying the natural exponential function to the sum of the scale estimate presented in the model results (see Table 5.18) to the sum of products between the vector Z and the elements of the vector β for the corresponding coefficient estimates for each of the patients characteristics in Z.

For example, a patient with the individual characteristics show in Table 5.19 would have the following moments for the Weibull distribution describing her survival function:

$$\boldsymbol{shape} = \exp(0.127) = \mathbf{1.1354}$$

$$\boldsymbol{scale} = \exp(7.516 + 0 \times 0.626 + 1.180 \times 1.441 + (-0.643) \times (-0.303) + (-1.779) \times 0.902 + (-1.447) \times 0.441 + (-0.918) \times (-0.341) + 0 \times (-0.121) + 1 \times (-0.460) + 0 \times (-0.478) + 0 \times (-0.260) + 0 \times 0.130 + 0 \times (-0.268) + 0 \times (-0.119) + 0 \times 0.251 + 1 \times (-0.438) + 0 \times (-0.105) + 0.563 \times (-0.411) + (-0.154) \times (-1.332) = \exp(6.557) = \mathbf{704.1561}$$

Table 5.19 – Individual patient characteristics of a random patient

Covariate	Value	Definition
intervention	0	EWS = 1; **UC = 0**
scale(ejection.fraction)	1.180	Ejection fraction (%); normalised
scale(age)	-0.643	Age (years); normalised
scale(sbp)	-1.779	Systolic blood pressure (mmHg) ; normalised
scale(bmi)	-1.447	Body mass index calculated as weight/height^2 (kg/m^2) ; normalised
scale(creatinine)	-0.918	Serum creatinine (µmol/l) ; normalised
(nyha.class)2	0	NYHA class 2 (Yes= 1; **No = 0**)
(nyha.class)3	1	NYHA class 3 (**Yes= 1**; No = 0)
(nyha.class)4	0	NYHA class 4 (Yes= 1; **No = 0**)
gender	0	Male = 1, **Female = 0**
smoker	0	Smoker = 1; **Non-smoker = 0**
diabetes	0	Diabetic = 1; **Non-diabetic = 0**
copd	0	COPD present = 1; **No COPD = 0**
recent.diagnosis	0	Diagnosis < 18 months = 1; **Diagnosis > 18 months = 0**
no.beta.blocker	1	**Without beta-blocker medication = 1**; On beta-blocker medication = 0
no.ace	0	Without ACE inhibitor medication = 1; **On ACE inhibitor medication = 0**
scale(age.ef)	0.563	Interaction between age and ejection fraction (product of variables above); normalised
scale(sbp.ef)	-0.154	Interaction between systolic blood pressure and ejection fraction (product of variables above); normalised

All continuous variables are normalised, in order to be in the same scale as covariate estimates

Please note that the procedure described above can be used for calculating the parameters for any of the distributions used in the survival analyses. However, there are crucial things to remember when calculating these parameters:

- The covariates are only used to model the location parameter (underlined in Table 5.17). The values for the covariates should only be used for calculating these parameters.
- All parameters bound to be strictly positive (see column "Parameters" in Table 5.17) are estimated on the log scale. In order to find the true value for the parameter that is used by the R functions, the estimates from flexsurvreg() must be converted using the natural exponential function.
- Parameters that can take any real value do not need to be transformed in any way. For instance, when using the log-normal distribution, the meanlog parameter is by adding the coefficient found for meanlog to the sum of products between the patient characteristics and the corresponding coefficient estimates for each of those characteristics.

e. Estimating time-to-event

The graphical representation of the survival function using a Weibull distribution with parameters $shape = 1.1354$ and $scale = 704.1561$ – found in the example from the previous section – is plotted through the code *curve((1-pweibull(x, shape=1.1354, scale = 704.1561)), xlim = c(0,3652.5)*, where *xlim* is the argument which determines that the x-axis, representing time in days, will have an interval between 0 and 3652.5 days (10 years). The resulting plot is shown in Figure 5.8. Note this represents the survival function of a patient with the characteristics shown in Table 5.19.

Time-to-event is estimated by replacing $S(t)$ in the survival equation (see column "CDF" in Table 5.17) by a random number drawn from a uniform distribution over the interval $[0,1]\big(U(0;1)\big)$ and solving it for t, where t is time-to-death. R easily does it through using the code *qweibull(runif (1), shape = 1.1354, scale = 704.1561)*, where *runif(1)* gives the random number between 0 and 1 that should replace $S(t)$ in the survival equation and *qweibull* computes the quantile function of the Weibull distribution.

The example where $S(t) = 0.35$ is shown graphically in dash in Figure 5.8. The intersection with the survival curve and the projection on the x-axis corresponds to the estimated time-to-death (734.96 days). The exact time-to-death is given by code *qweibull(1-0.35, shape = 1.1354, scale = 704.1561)*.

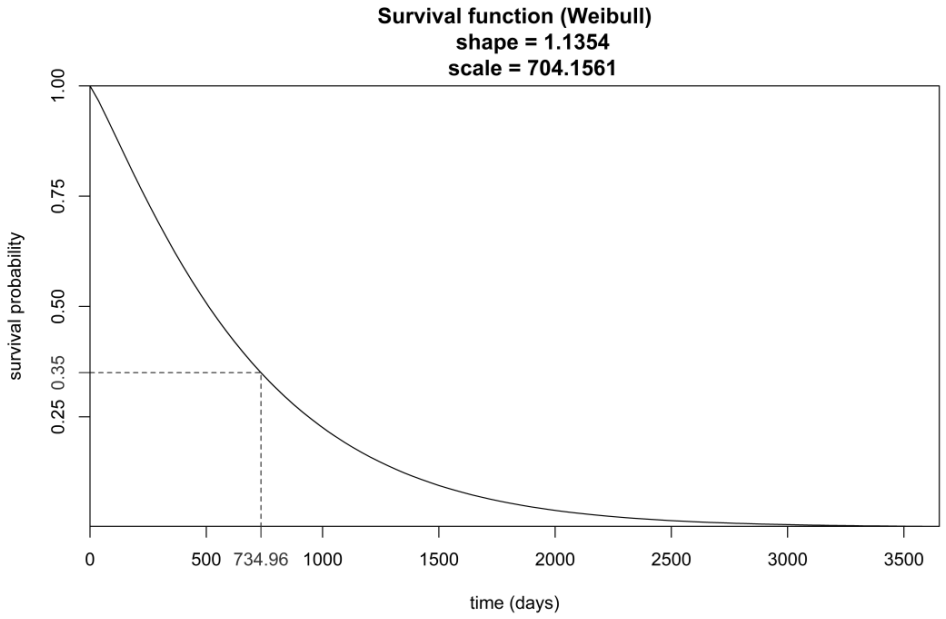


Figure 5.8 – Plot of the survival function using the Weibull distribution with parameters set by the user

It is worth mentioning that time-to-death calculations can be performed for any of the other discussed distributions, simply by replacing *pweibull* and *qweibull* with the corresponding functions for the parametric distributions as per Table 5.16, and by using the respective arguments for the parameters in those functions (see column "Parameters" in Table 5.17).

f. Probabilistic sensitivity analysis

Probabilistic sensitivity analysis was implemented as a double loop: an inner loop, in which a number of patients are sampled with replacement from the baseline population, and an outer loop, in which values of the input parameters of the model are randomly drawn.

For each of the outer loop runs, it is necessary to consider the variability observed in the parametric regression models, for the parameters estimates and for the covariate coefficient estimates.

All the steps for estimating time-to event are followed as described above, except for the numeric vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, where $\beta_1, \beta_2, \dots, \beta_n$ are the estimates for the parameters and the coefficients of the covariates in the model. The vector β is extracted by the function *mvtnorm::rmvnorm(1,coef(surv.model),vcov(surv.model)),* where *surv.model* is a *flexsurvreg* object representing the survival model used in the calculations. This function belongs to the *mvtnorm* package and provides estimates for

a multivariate normal distribution taking into account the covariance matrix of the parameter and coefficient estimates.

Appendix 5.1-B – Time-to-hospitalisation

This appendix shows the survival analyses results for time-to-hospitalisation. As stated in the beginning of Appendix 5.1, the detailed description of the performed analyses can be found in Appendix 5.1-A.

It is also worth mentioning that hospitalisation – the event of interest in these survival analyses – refers to the first hospitalisation, while subsequent hospitalisations are assumed to follow the same distribution as the first hospitalisation.

1. Kaplan-Meier estimates

The KM estimate for hospitalisation for the usual care (UC) and the early warning system (EWS) interventions is presented in Figure 5.9.

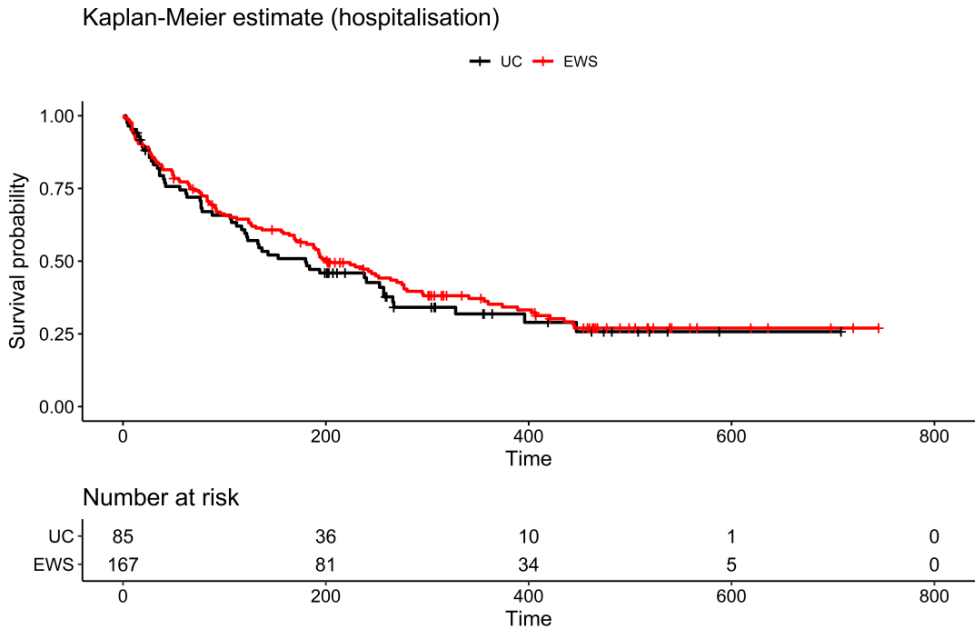


Figure 5.9 – Kaplan-Meier estimates for hospitalisation

2. Survival curves difference

a. log-rank or Mantel-Haenszel test

Table 5.20 presents the results of the log-rank or Mantel-Haenszel test for hospitalisation.

Table 5.20 – log-rank or Mantel-Haenszel test for hospitalisation

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
UC	85	54	50.2	0.283	0.412
EWS	167	109	112.8	0.126	0.412
Chisq = 0.4 on 1 degrees of freedom, p = 0.5					

The results do not show a statistically significant difference between the survival curve of UC and EWS, suggesting that there is no difference in hospitalisation patterns between the interventions.

b. Peto & Peto modification of the Gehan-Wilcoxon test

Table 5.21 shows the results of the Peto & Peto modification of the Gehan-Wilcoxon test.

Table 5.21 – Peto & Peto modification of the Gehan-Wilcoxon test for hospitalisation

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
UC	85	37.3	34.3	0.257	0.513
EWS	167	71.6	74.6	0.118	0.513
Chisq = 0.5 on 1 degrees of freedom, p = 0.5					

The results do not show a statistically significant difference between the survival curve of UC and EWS, suggesting that there is no difference in hospitalisation patterns between the interventions.

3. Cox proportional hazards model

a. Parameter estimation

We fist estimated the hazard ratio between the interventions by fitting a Cox proportional hazards model where the intervention was the only covariate (see results in Table 5.22).

Table 5.22 – Cox proportional hazards regression model results for hospitalisation (intervention only)

covariate	exp(coefficient)	Standard error	P-value
intervention	0.899	0.167	0.523
n = 252, number of events = 163 Likelihood ratio test = 0.4 on 1 df, p = 0.5255			

The results of this model show a hazard ratio between the EWS and UC of 0.899. In other words, a patient who is treated with the EWS has a 10.1% decrease on the hazard of being hospitalised when compared to the UC intervention. The hazard ratio, however, is not statistically significant.

We subsequently estimated the impact of multiple covariates in survival by running another regression model. The definition of the covariates used in the Cox proportional hazards model is presented in Table 5.10. The results of the regression model for hospitalisation are shown in Table 5.23.

Table 5.23 – Cox proportional hazards regression model results for hospitalisation

covariate	exp(coefficient)	Standard error	P-value
intervention	0.964	0.176	0.834
ejection.fraction	3.436	0.635	0.052
age	1.894	0.325	0.05
sbp	1.063	0.265	0.817
bmi	1.077	0.09	0.411
creatinine	1.273	0.08	0.003
(nyha.class)2	1.345	0.234	0.206
(nyha.class)3	1.567	0.249	0.072
(nyha.class)4	2.253	0.373	0.029
gender	1.319	0.224	0.216
smoker	1.063	0.278	0.827
diabetes	1.02	0.175	0.908
copd	1.126	0.189	0.53
recent.diagnosis	0.8	0.174	0.199
no.beta.blocker	1.713	0.176	0.002
no.ace	0.691	0.225	0.1
age.ef	0.275	0.622	0.038
sbp.ef	0.855	0.572	0.784

n = 252, number of events = 163
Concordance= 0.64 (se = 0.022)
Likelihood ratio test = 40.56 on 18 df, p = 0.002
Wald test = 40.54 on 18 df, p = 0.002
Score (logrank) test = 42.11 on 18 df, p = 0.002

In our model, there was a statistically significantly association between older patients, patients who were not on beta blocker treatment, and patients with higher levels of creatinine and an increased chance for being hospitalised. Although being treated with the EWS reduced the chance of being hospitalised when compared to the UC intervention, this relationship was not statistically significant.

b. Proportional hazards assumption test

The results of the *cox.zph()* function for our models are shown in Table 5.24 and Table 5.25, for intervention as the only covariate and the full set of covariates, respectively.

Table 5.24 – Schoenfeld residuals of the Cox proportional hazards model covariates for hospitalisation (intervention only)

covariate	chisq	df	P-value
factor(intervention)	0.12	1	0.73
GLOBAL	0.12	1	0.73

Table 5.25 – Schoenfeld residuals of the Cox proportional hazards model covariates for hospitalisation

covariate	chisq	df	P-value
factor(intervention)	0.170476102	1	0.679689027
scale(ejection.fraction)	7.180377709	1	0.007370519
scale(age)	0.025177383	1	0.873925862
scale(sbp)	0.20188116	1	0.653206694
scale(bmi)	0.246169044	1	0.619785583
scale(creatinine)	0.464406238	1	0.49557188
factor(nyha.class)	2.283562114	3	0.515677282
factor(gender)	1.013978369	1	0.313951628
factor(smoker)	0.622900127	1	0.429971618
factor(diabetes)	1.316978403	1	0.251135054
factor(copd)	0.87328747	1	0.350046803
factor(recent.diagnosis)	0.483061493	1	0.487038849
factor(no.beta.blocker)	0.53585546	1	0.464155452
factor(no.ace)	2.078203506	1	0.149415939
scale(age.ef)	5.292681555	1	0.02141522
scale(sbp.ef)	5.234398519	1	0.022144464
GLOBAL	19.01507495	18	0.390895718

From the results in Table 5.25 we can observe that the test is not statistically significant for most of the covariates and that the global test is also not statistically significant. A lack of statistical significance is also observed in the test for the model with the intervention being the only covariate (Table 5.24). We can therefore assume that the proportional hazards assumption holds.

c. Influential observations test

Figure 5.10 and Figure 5.11 show the results of the influential observations test for our models, for intervention as the only covariate and the full set of covariates, respectively.

The above index plots show that comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that most of the observations are not much influential individually, even though some outliers (larger dfbeta values) can be spotted for some covariates in the model with the full set of covariates (Figure 5.11).

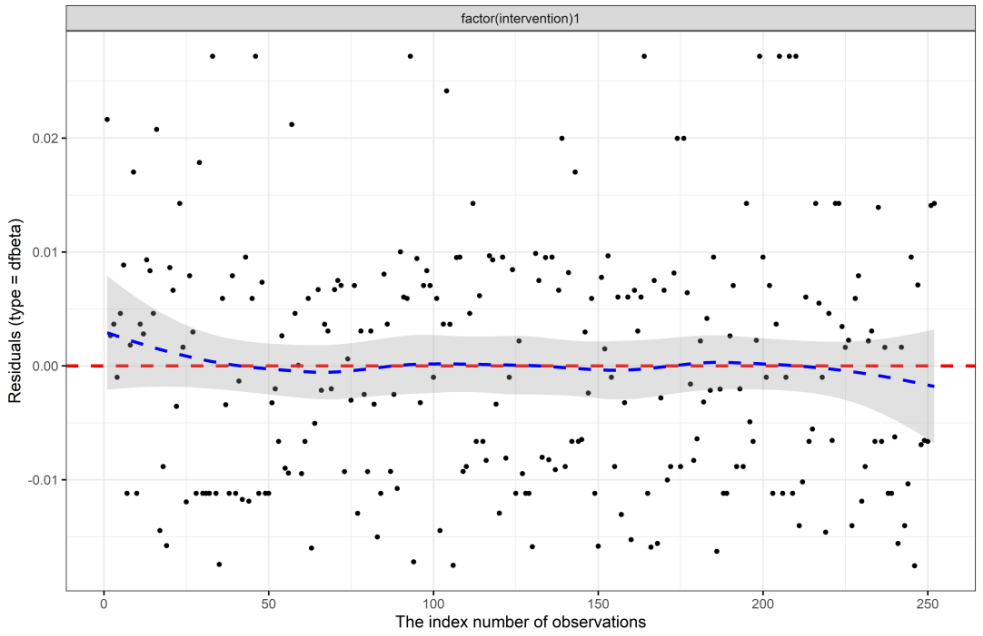


Figure 5.10 – dfbeta residuals for the Cox proportional hazards model for hospitalisation with intervention as the only covariate

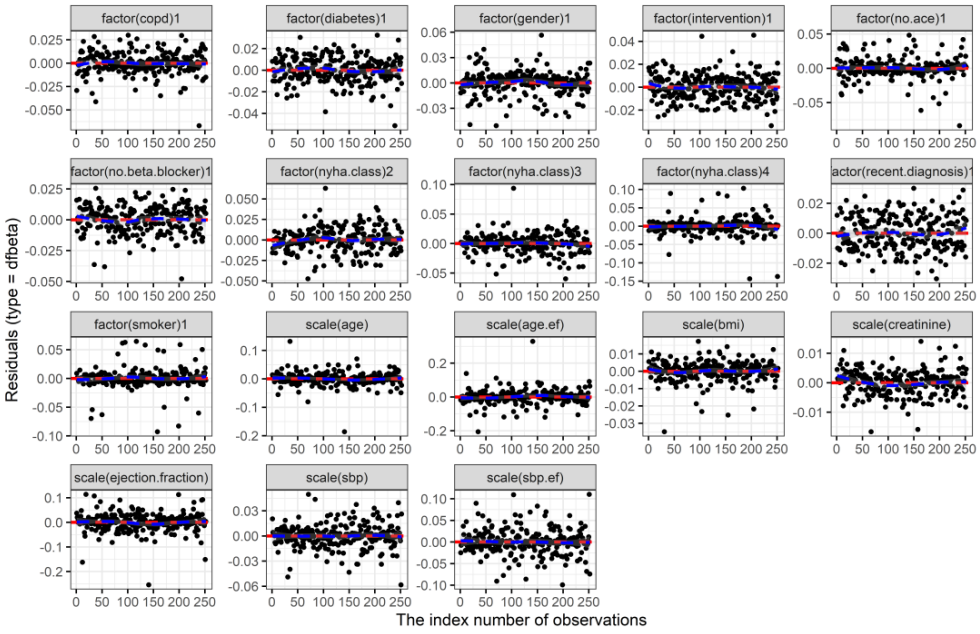


Figure 5.11 – dfbeta residuals for the Cox proportional hazards model covariates for hospitalisation

d. Log cumulative hazards

Figure 5.12 shows an illustration of a log cumulative hazard plot for the Kaplan Meier curves for hospitalisation previously shown in Figure 5.9. The graph demonstrates that there are no seemingly changes in the hazard during the trial duration and that hazards are reasonably proportional between the two treatment groups. This indicates that a single parametric model may be suitable to model survival, including an exponential model.

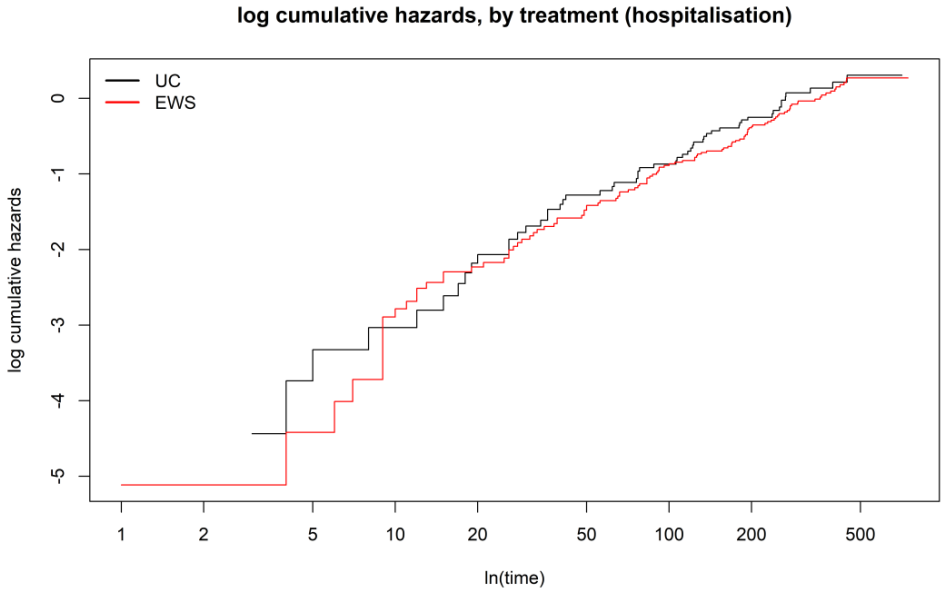


Figure 5.12 – log cumulative hazards, by treatment (death)

4. Parametric survival modelling

a. Visual inspection

The graphical representation of the fitting of the exponential, Weibull, Gompertz, log-logistic, log-normal, and generalised gamma models to time-to-hospitalisation data is shown in Figure 5.13 and Figure 5.14. Although we are using the intervention as a covariate and assuming proportional hazards, we are showing the graphs separately for the UC and for the EWS (Figure 5.13 and Figure 5.14, respectively), for the purpose of clarity.

The visual inspection analysis suggests a worse fit of the exponential and the Weibull models to the data, as they do not seem to follow the noticeable flattening of the Kaplan-Meier curves at the end of trial follow-up.

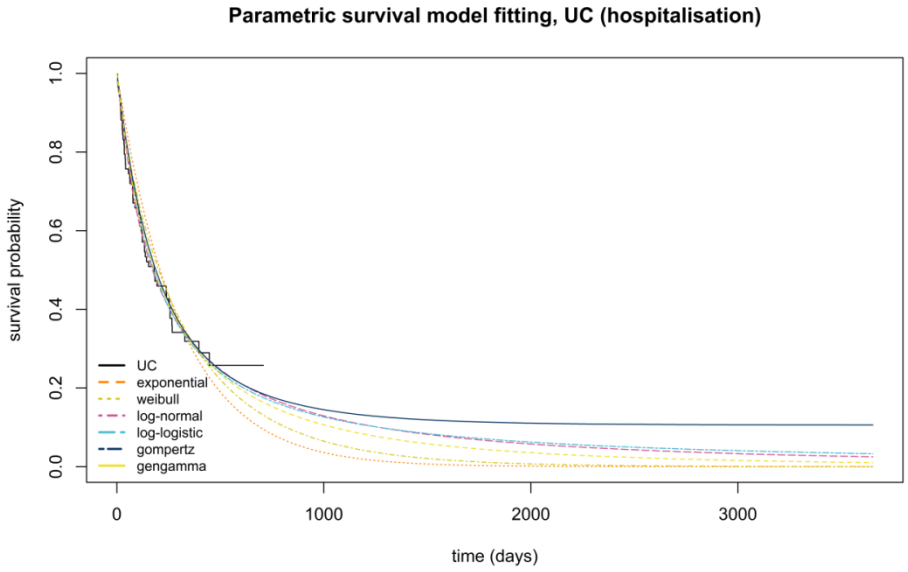


Figure 5.13 – Fitting of parametric survival models to time-to-hospitalisation (UC)

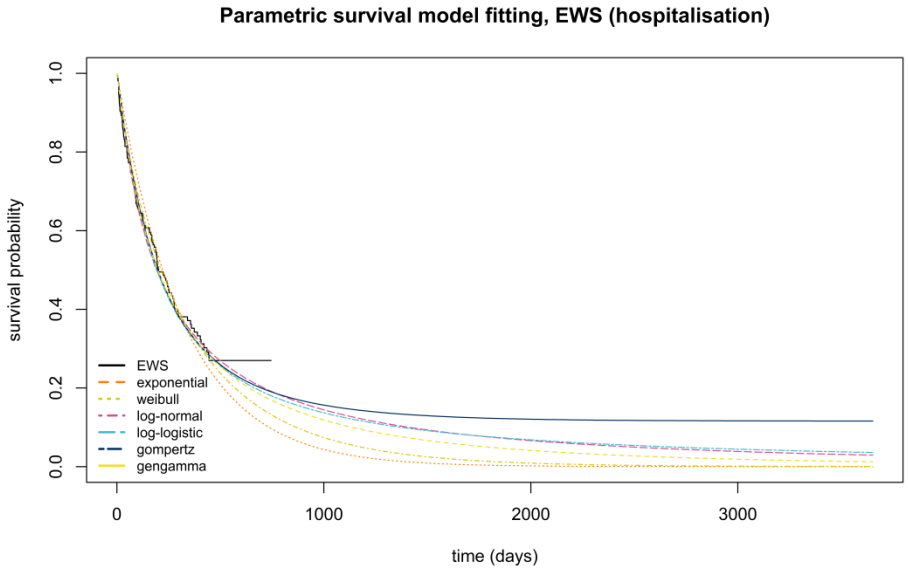


Figure 5.14 – Fitting of parametric survival models to time-to-hospitalisation (EWS)

b. Goodness-of-fit to the observed data

The results of the goodness-of-fit tests for time-to-hospitalisation using the AIC and BIC are presented in Table 5.26.

Table 5.26 – Goodness-of-fit tests for parametric survival modelling of hospitalisation

Parametric model	AIC	BIC
Exponential	2194.65	2261.71
Weibull	2190.86	2261.45
Log-normal	2187.67	2258.26
Log-logistic	2188.74	2259.33
Gompertz	2187.52	2258.11
Generalised gamma	2188.61	2262.73

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion.

As suggested by the visual inspection, the higher values for the AIC and the BIC corroborate the worse fit of the exponential and the Weibull models to the observed data. The higher value for the generalised gamma in the BIC can be explained by the extra parameter used for defining this distribution.

c. Clinical validity and external data

The model using the Gompertz distribution reaches an early plateau (around that determines that approximately 20% of the patients are not hospitalised, which seems very unlikely. If a patient is not hospitalised, it should be because of the competing risk of dying and not because of the extrapolation of the survival curve.

We could not find in the literature information for the first hospitalisation, as this would imply having access to patient-level data, which should come from a very similar population in order to be comparable.

Since we used the time to first hospitalisation and we assumed that the time to a subsequent hospitalisation follows the same distribution – as subsetting the data for the patients who experienced two or more hospitalisations would lead to very low sample sizes –, we used the hospitalisation rate found in the trial and compared it to the hospitalisation rates found in the model for simulations ran for the same set of 1000 patients using each of the parametric distributions used for the extrapolation of time-to-hospital and the Weibull distribution as the parametric distribution for time-to-death. For this exercise, we set the time-to-outpatient visit to 1 year, as we would like to have a periodic update of the patient characteristics. The hospitalisation rate per follow-up year found in the trial for was 1.254 for UC and 1.218 for EWS; the results from the simulations can be found in Table 5.27.

Table 5.27 – Hospitalisation rates from simulations with different parametric distributions

Parametric model	Hospitalisation rate		Difference from trial	
	UC	EWS	UC	EWS
Exponential	1.270	1.266	1.23%	3.98%
Weibull	1.282	1.288	2.20%	5.75%
Log-normal	1.370	1.312	9.24%	7.75%
Log-logistic	1.337	1.296	6.59%	6.46%
Gompertz	1.299	1.326	3.53%	8.88%
Generalised gamma	1.349	1.311	7.56%	7.70%

d. Conclusion

Considering the relatively small differences between the hospitalisation rates found in the simulations and the ones found in the trial, we relied on the considerations made in the visual inspection and the goodness-of-fit to the observed data for choosing the log-normal for the base-case analysis in our model, as it has the lower AIC and BIC values (after the Gompertz). However, in order to analyse the impact of the choice of different parametric models on the results of the cost-effectiveness model, sensitivity analyses should be used.

Appendix 5.2 – Generalised Linear Model (logistic regression) for calculating the probability of dying in hospital

A Generalised Linear Model (GLM) is a general type of regression model that can be applied to a very wide variety of response variables. GLMs are important because they provide a natural way to analyse certain types of data such as discrete data that might not be modelled well using linear regression. GLMs differ from linear models in three ways:

- The distribution of the response is modelled using a distribution other than the normal distribution.
- The variance of the response is not assumed to be constant but can vary as a function of the mean response.
- The relationship between the mean response and the predictor variables is allowed to be non-linear.

A GLM consists of two: (1) the link function and (2) the variance function. The link function describes how the mean response is modelled as a function of the explanatory variables, while the variance function specifies a certain form for the variances of the response variable.

The link function is nothing more than a transformation applied to the mean response on the left side of the regression model equation. In general, a GLM specifies the relationship between the mean response and explanatory variables (covariates) as: $g(\mu) = \beta_0 + \beta_1 Z_1 + \cdots + \beta_n Z_n$, where g is the link function, μ the mean response, the β_0 is the intercept coefficient, and $\beta_1, \ldots, \beta_n$ the coefficients for the covariables, which are represented by $Z_1, \ldots, Z_n$. Please note that the right side of the equation is the same for all regression models.

One of the most commonly used GLMs is the logistic regression model, which is characterized by the following features:

- The response variable Y is a binary indicator variable (takes on values 0 or 1).
- The relationship between the mean and the variance of the response has the form: $var(Y) = p(1-p)$, where $p = P(Y = 1)$.
- The relationship between p and the predictor variables is based on the log-odds (also called logit) transformation: $log\frac{p}{1-p} = x$.

The logistic regression model is used when the quantification of a treatment effect measure is done with an odds ratio, and it is widely used because the fitted values will always lie in the interval $[0; 1]$, which is extremely handy when we are dealing with probabilities.

To show the fact above, let $log\frac{p}{1-p} = x$. If so, then $= \frac{e^x}{1+e^x}$, which lies in (0,1) for all values of x (see Figure 5.15).

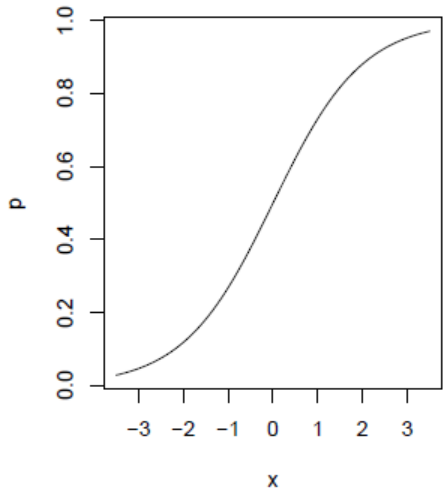


Figure 5.15 – Probability (p) as a function of the log-odds (x)

In order to simulate the probability of dying in hospital, a logistic regression model was used to estimate the log-odds for having the outcome (dying in hospital) depending on patient characteristics. The patient characteristics are the covariates used to explain the outcome (represented by $\beta_0 + \beta_1 Z_1 + \cdots + \beta_n Z_n$ above).

Table 5.28 presents the covariates used for explaining death in hospital. They were chosen because they represent the individual patient characteristics that were considered to have an effect on the probability of a given patient dying in hospital. Either because of their impact on the heart functionality (myocardial infarction and chronic atrial fibrillation), or because they are associated with a decrease in general health status of the patient – linked with the pathophysiology of the underlying disease (age, gender, and number of previous hospitalisations) or comorbidities associated with described complications in heart failure patients (diabetes and chronic obstructive pulmonary disease).

Table 5.28 – Covariates used in the logistic regression for explaining death in hospital

Covariate	**Variable type**	**Definition**
age	Continuous	Age (years)
gender	Categorical	Male = 1, Female = 0
myocardial.infarction	Categorical	Previous history of myocardial infarction = 1; No previous history of myocardial infarction= 0
chronic.atrial.fibrillation	Categorical	Previous history of chronic atrial fibrillation = 1; No previous history of chronic atrial fibrillation = 0
diabetes	Categorical	Diabetes present = 1; No diabetes = 0
copd	Categorical	COPD present = 1; No COPD = 0
previous.hosp	Continuous	Number of previous hospitalisations simulated in the model

Categorical variables in R are referred to as factors. In the results of the model shown in the R console, the covariate intervention is labelled as *factor(gender)1*, meaning that the results for the coefficients are given for a value of 1 for the covariate gender (i.e. male patient).

For the purpose of simplifying the results shown, *factor* was omitted from the covariates.

The model used can therefore be formalised by the expression: $log\frac{p}{1-p} = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3 myocardial.infarction + \beta_4 chronic.atrial$fibrillation $+ \beta_5 diabetes + \beta_6 copd + \beta_7 previous.hosp$, where p is the probability of dying in hospital. The results of this model are presented in Table 5.29.

Table 5.29 – Results of the logistic regression for estimating the probability of dying in hospital

	Coefficient estimate	Standard error	P-value
(Intercept)	-3.822	0.383	< 0.001
age	0.022	0.15	0.077
gender	-0.059	0.345	0.864
myocardial.infarction	0.099	0.276	0.721
chronic.atrial.fibrillation	0.196	0.273	0.472
diabetes	0.143	0.26	0.582
copd	0.314	0.265	0.235
previous.hosp	0.290	0.253	0.251

Deviance Residuals

Min	1Q	Median	3Q	Max
-0.7689	-0.5723	-0.5084	-0.4223	2.4579

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 456.23 on 586 degrees of freedom
Residual deviance: 447.91 on 579 degrees of freedom
AIC: 463.91
Number of Fisher Scoring iterations: 5

The column "Coefficient estimate" represents the values of the different betas in the expression formalised above, and their exponentiated value can be interpreted as odds ratios.

For calculating the log-odds of a specific patient, we can use the coefficients and the individual characteristics of a patient. For example, the log-odds for patient A, who is a 60 year-old male without previous history of myocardial infarction and chronic atrial fibrillation, is diabetic, does not have COPD, and has not been hospitalised previously in the simulation are calculated as follows:

$$log\frac{p}{1-p} = -3.822 + 0.022 \times 60 - 0.059 + 0.099 \times 0 + 0.196 \times 0 + 0.143 \times 1 + 0.314 \times 0 + 0.290 \times 0 = -2.418$$

In order to obtain the probability of dying in hospital, we must covert the log-odds using the formula, $p = \frac{e^x}{1+e^x}$. Thus:

$$p = \frac{e^{-2.418}}{1+e^{-2.418}} = 0.0818$$

The probability of dying in hospital for patient A is 8.18%.

Please note that we can also get to this value in R through the code:

predict(p.death.hosp, newdata = data.frame(age = 60, gender = 1, myocardial.infarction = 0, chronic.atrial.fibrillation = 0, diabetes = 1, copd = 0, previous.hosp = 0), type="response")

0.08248722

The slight difference between the two calculations has to do with rounding, as R does not round when using the function *predict()* and we rounded our coefficients to the third decimal place in our calculations above.

As another example, please take patient B, who is 85 year-old female with previous history of myocardial infarction and chronic atrial fibrillation. She is diabetic, has COPD, and has not been hospitalised previously in the simulation. Her probability of dying in hospital is 23.37%, as shown by the result of the function *predict()* in R:

predict(p.death.hosp, newdata = data.frame(age = 85, gender = 0, myocardial.infarction = 1, chronic.atrial.fibrillation = 1, diabetes = 1, copd = 1, previous.hosp = 0), type="response")

0.2336884.

The incorporation of the results of the logistic regression in the model is done through the following steps:

1. Every time the model is processing the event hospitalisation, the patient characteristics at the time of simulation are extracted.
2. The probability of dying in hospital is calculated for that patient using the characteristics extracted in 1 (see examples above).
3. A random number is drawn from a uniform distribution over the interval $[0, 1]$ using the function runif().
4. If the number generated in 3 is lower than the probability of dying in hospital found in 2, the simulated patient dies and death is the next processed event; if it is higher, the simulation continues as per the model structure. Using the examples above, if the number drawn in 3 would be 0.156543, patient A would not die in the simulation, unlike patient B, who would die, as $0.156543 < 0.2336884$.

Online Appendix 5.1 – Assessment of the Validation Status of Health-Economic decision models tool (AdViSHE)

Available at:

https://drive.google.com/file/d/1GKsjtNznsLacTYS-xgHgF_t2rLDiaBqu/view?usp=sharing

Chapter 6

Cost-effectiveness of a home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands

Fernando Albuquerque de Almeida, Isaac Corro Ramos, Maiwenn Al and Maureen Rutten-van Mölken

JMIR mHealth and uHealth (In review)

Abstract

Background: Heart failure (HF) is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. Home telemonitoring (HTM) facilitates frequent or continuous assessment of disease signs and symptoms, while it has been shown to improve compliance by involving patients in their own care and to prevent emergency admissions by facilitating early detection of clinically significant changes. Diagnostic algorithms (DAs) are predictive mathematical relationships that make use of a wide range of collected data for calculating the likelihood of a particular event happening and utilise this output for prioritising patients with regards to their treatment.

Objective: Assessing the cost-effectiveness of HTM and a DA in the management of heart failure in the Netherlands. Three interventions were analysed: usual care (UC), HTM, and HTM+DA.

Methods: A previously published discrete event simulation model was used. The base-case analysis was performed according to the Dutch guidelines for economic evaluation. Sensitivity, scenario, and value of information analyses were performed. Particular attention was given to the cost-effectiveness of the DA at various levels of diagnostic accuracy of event prediction and to different patient subgroups.

Results: HTM+DA extendedly dominates HTM and it has a deterministic incremental cost-effectiveness ratio versus UC of €27,712 per quality-adjusted life year (QALY). The model showed robustness in the sensitivity and scenario analyses. HTM+DA had a 96.0% probability of being cost-effective at a €80,000/QALY threshold. An optimal point for the threshold value for the alarm of the DA in terms of its cost-effectiveness was estimated. NYHA class IV patients were the subgroup with the worst cost-effectiveness results versus UC, while HTM+DA was found to be the most cost-effective for patients <65 years-old and for patients in NYHA class I.

Conclusions: Although increased costs of adopting HTM and DA in the management of HF may seemingly be an additional strain on scarce health care resources, the results of this study demonstrate that, by increasing patient life expectancy by 1.28 years and reducing their hospitalisation rate by 23% when compared to UC, the use of these technologies may be seen as an investment, as HTM+DA extendedly dominates HTM and is cost-effective versus UC at normally accepted thresholds in the Netherlands.

Introduction

Heart failure (HF) is a major health concern associated with significant morbidity, mortality, and reduced quality of life for patients. Approximately 1–2% of the adult population in the West has HF and its prevalence rises above 10% for the population with 70 years of age and older (16). In 2019, the Dutch prevalence of HF was estimated to be 238,700, with an incidence of 37,400 new cases and a total of 7,264 deaths due to HF (233). Accordingly, HF is responsible for elevated health care costs in the Netherlands: €817 million in 2017, corresponding to 8% of the costs for cardiovascular diseases and around 1% of the total health care expenditure for that year (233). From the total HF costs, 45% are attributable to care provided in the hospital and 43% are spent on care for the elderly (long-term institutional elderly care, assisted-living facilities for the elderly, and home care.) (233).

Remote patient monitoring is a patient management approach defined as the use of information and communication technologies to monitor and transmit physiological data related to patient health status between geographically separated individuals (234). Home telemonitoring (HTM) is the particular case in which the monitoring and the transmission of data are done from the patients' home. HTM facilitates frequent or continuous assessment of disease signs and symptoms, while it has been shown to improve compliance by involving patients in their own care and to prevent emergency admissions by facilitating early detection of clinically significant changes (235). The use of information and communication technologies in the management of chronic diseases has become increasingly important, especially since the COVID-19 pandemic, when routine care had to be postponed or replaced by remote alternatives. There is evidence showing that HTM can have a positive impact on both mortality and hospital admissions (101, 102, 236), while other studies question the effectiveness (103) and cost-effectiveness (237) of home-based monitoring systems.

Diagnostic algorithms (DAs) can be defined as predictive mathematical relationships that make use of a wide range of collected data for calculating the likelihood of a particular event happening (e.g., death or hospitalisation). DAs use this output for prioritising patients with regards to their treatment and through raising alarms that trigger follow-up actions if the probability of having the event exceeds a pre-defined threshold. Evidence shows that data-driven approaches looking at trends and patterns of change in recorded parameters improve the accuracy of detecting disease deterioration when compared to clinical decision rules (171, 173, 225, 226). Coupled with the fact that large number of parameters associated with events in HF can be measured with HTM, it is expected that advanced algorithms with better diagnostic performance will result in time efficiencies in analysing the data generated with HTM systems and, in that way, improve clinical decision making through raising alerts in a

manner that can be intuitively used by clinicians with a high degree of confidence (174). However, health care funds are limited and we must deal with the decision of allocating scare resources to patient subgroups for which new interventions are most beneficial.

The objective of this study was to assess the cost-effectiveness of HTM and a DA in the management of HF in the Netherlands. A base-case analysis was performed and structural and parametric uncertainty was assessed through scenario, sensitivity, and value of information (VOI) analyses. Further, we focused particularly on the assessment of the cost-effectiveness of the DA at different levels of diagnostic accuracy of event prediction, i.e. different points of its receiver operating characteristic (ROC) curve, as well as on the cost-effectiveness of the interventions under analysis for a wide range of patient subgroups.

Methods

Model structure

The patient-level discrete event simulation (DES) model used for the analysis was developed and described in detail elsewhere (238). Unlike other published health economic models for HF, this is a singular model that includes a wide range of patient characteristics and outcomes. The model consists of a series of regression equations describing the statistical associations between the patient characteristics and changes in intermediate and final outcomes over time. The time-to-event regression equations were estimated using the patient-level data of the Trans-European Network – Home-Care Management System (TEN-HMS) study (126). The model simulates the time to an outpatient visit, hospitalisation and death. Intermediate outcomes generated from the model are the number of outpatient visits, hospitalisations, and avoided hospitalisations. Final outcomes are the total life years, quality-adjusted life years (QALYs), and costs.

Each patient is simulated for the three interventions included in the cost-effectiveness analysis: (i) usual care (UC) – patient management plan implemented by the patient's primary care physician (126), (ii) HTM (as described in the TEN-HMS original publication (126)), and (iii) HTM with the addition of a DA (HTM+DA).

Model population

In the base-case analysis, patients are randomly sampled (with replacement) from the entire population included in the TEN-HMS study (126). The baseline patient and disease characteristics of the model population are shown in Table 6.1. This patient population is assumed to be representative of the Dutch HF patient population.

Table 6.1 – Baseline patient and disease characteristics of the model population

	Baseline characteristics of the starting population
Sample size	426
Ejection fraction (EF), % (mean)	25.06
Age, years (mean)	67.56
Systolic blood pressure (SBP), mm Hg (mean)	114.24
Body mass index (BMI), kg/m2 (mean)	26.17
Creatinine, μmol/l (mean)	135.71
NYHA class 1, %	18.5
NYHA class 2, %	43.4

NYHA class 3, %	31.0
NYHA class 4, %	7.1
Gender (male), %	77.5
Smoker, %	12.2
Diabetes, %	35.0
Chronic obstructive pulmonary disease (COPD), %	24.4
Recent diagnosis, %	43.9
No beta-blocker medication, %	37.3
No ACE inhibitor medication, %	18.5
Myocardial infarction, %	56.8
Chronic atrial fibrillation, %	26.3

Abbreviations: ACE, angiotensin-converting enzyme; BMI, body mass index; COPD, chronic obstructive pulmonary disease; EF, ejection fraction; NYHA, New York Heart Association; SBP, systolic blood pressure.

Dutch cost-effectiveness threshold

The cost-effectiveness threshold in the Netherlands depends on the burden of disease as measured by the fraction of QALYs that people lose relative to the situation in which the disease had been absent (proportional shortfall) (239-241). The appropriate cost-effectiveness threshold, which represents the societal willingness-to-pay (WTP) for an additional QALY for that specific patient population, can be calculated using the Institute for Medical Technology Assessment Disease Burden Calculator (242).

Base-case analysis

The base-case analysis was conducted in accordance with the Dutch guidelines for economic evaluations in healthcare (210). A societal perspective was adopted, which considered costs including all costs inside the healthcare sector, patient and family, and other sectors, regardless of who is paying for those costs, productivity losses assessed using the friction cost method (81), and future unrelated medical costs. All costs are reported in 2020 euros; where 2020 figures were not available, older costs were inflated using the general price index from the Dutch Central Bureau of Statistics (243). Health outcomes (effects) were presented in life years and QALYs and discounted at 4.0%, while costs were discounted at 1.5%. The analysis adopted a lifetime time horizon and the model was run for 1,000 patients. An overview of model

input parameters is presented in Table 6.2 and explained in detail in the following sections.

Treatment effect of HTM (compared to UC)

When compared to UC, HTM is modelled to increase time-to-hospitalisation and time-to-death, while decreasing time-to-outpatient visit.

The treatment effect of HTM on time-to- hospitalisation and time-to-death is modelled through the use of parametric models (exponential, Weibull, log-normal, log-logistic, Gompertz, and generalised gamma) fitted to empirical time-to- hospitalisation and time-to-death data (Kaplan-Meier curves) for HTM and UC from the TEN-HMS trial (126). The models assumed proportional hazards between HTM and UC. In the base-case analysis, a Weibull distribution was used for extrapolating time-to-death and a log-normal distribution for extrapolating time-to-hospitalisation. The distributions were chosen according to the recommendations issued by the Decision Support Unit commissioned by The National Institute for Health and Clinical Excellence (206). The details of the survival analysis can be found in the original publication of the model (238).

When a patient is hospitalised, there is a chance of dying in the hospital. For predicting it, we ran a logistic regression where the probability of dying in hospital is explained by age, gender, previous history of myocardial infarction and/or chronic atrial fibrillation, comorbidities (diabetes and/or COPD), and the number of previous hospitalisations.

Time-to-outpatient visit is a parameter set by the user in the model. Since there is no periodic outpatient visit suggested in Dutch or international guidelines – as it is recommended that time to the next consultation is scheduled by the accompanying physician and based on the clinical status of the patient (11, 12) –, we assumed that the time-to-outpatient visit for the population under analysis is properly represented by the observations in the TEN-HMS study (126): 2.81 months for UC and 1.69 months for HTM-based interventions. This assumption is strengthened by the fact that 161 out of the 426 patients (37.8%) included in the TEN-HMS trial were treated in Dutch hospitals (126).

Table 6.2 – Model input parameters

Parameter (source)		Probabilistic sensitivity analysis		Deterministic sensitivity analysis (95% confidence interval)		
	Mean value	SE	Distribution	Lower bound	Upper bound	Observations
Model settings						
Discount rate (costs) (210)	4.0%	N/A	N/A	0%	8%	Dutch EE guidelines
Discount rate (effects) (210)	1.5%	N/A	N/A	0%	3%	Dutch EE guideline
Time horizon (210)	Lifetime	N/A	N/A	N/A	N/A	Dutch EE guidelines
Treatment effect						
Time-to-death (distribution) (238)	Weibull	N/A	N/A	N/A	N/A	Uncertainty assessed in the scenario analyses
Time-to-hospitalisation (distribution) (238)	Log-normal	N/A	N/A	N/A	N/A	Uncertainty assessed in the scenario analyses
Time-to-outpatient visit (UC, months) (126)	2.81	10% of the mean	Normal	2.46	3.13	None
Time-to-outpatient visit (HTM, months) (126)	1.69	10% of the mean	Normal	1.59	1.79	None
Diagnostic algorithm						
Sensitivity (244)	0.52	N/A	N/A	N/A	N/A	Uncertainty assessed in the scenario analyses for the DA
False positive rate (244)	0.03	N/A	N/A	N/A	N/A	Uncertainty assessed in the scenario analyses for the DA
Proportion avoidable hospitalisations (208)	50%	20% of the mean	Normal	33.6%	66.4%	None
Costs						
Outpatient visit (UC) (126, 211)	€44.50	20% of the mean	Gamma	€30.94	€60.08	None
Outpatient visit (HTM) (126, 211)	€43.30	20% of the mean	Gamma	€30.11	€58.46	None

Other HF-related care provider contacts (UC) (126, 211)	€188.38	20% of the mean	Gamma	€130.98	€254.33	None
Other HF-related care provider contacts (HTM) (126, 211)	€623.61	20% of the mean	Gamma	€433.59	€841.93	None
Hospitalisation (126, 211, 233)	€4,404.46	20% of the mean	Gamma	€3,062.36	€5,946.44	None
HTM device (per year) (205)	€1,257.75	20% of the mean	Gamma	€1,059.87	€1,469.69	None
Managing alarm (211)	€18.38	20% of the mean	Gamma	€12.78	€24.81	None
Drug costs (per year) (126, 245)	€286.44	20% of the mean	Gamma	€199.16	€386.72	None
Travelling expenses (outpatient visit) (126, 211, 246)	€3.75	20% of the mean	Gamma	€2.61	€5.06	None
Travelling expenses (hospitalisation) (211, 246)	€4.68	20% of the mean	Gamma	€3.25	€6.32	None
Informal care (per year) (126, 211, 243)	€2,098.28	20% of the mean	Gamma	€1,458.90	€2,832.88	None
Utilities						
NYHA class I (205)	0.87976	0.00827	Beta	0.86588	0.89308	None
NYHA class II (205)	0.71178	0.00944	Beta	0.69615	0.72720	None
NYHA class III (205)	0.61405	0.01349	Beta	0.59176	0.63614	None
NYHA class IV (205)	0.49228	0.03032	Beta	0.44243	0.54220	None
Utility multiplier (outpatient visit)	1	N/A	N/A	N/A	N/A	Assumption; excluded from uncertainty analyses*
Utility multiplier (hospitalisation) (207)	0.82	10% of the mean	Normal	0.69	0.95	None

Abbreviations: EE, economic evaluation; HTM, home telemonitoring; LY, life year; N/A, not applicable; NYHA, New York Heart Association; QALY, quality-adjusted life year; SE, standard error; UC, usual care.

* Depending on the rate of outpatient visits, positive values may generate higher QALYs when compared to LYs.

Treatment effect of the DA (when added to HTM alone)

To model the treatment effect of adding the DA to HTM we considered the DA as a binary test for predicting a hospitalisation. Depending on the threshold value for the alarm of the DA to be raised, the DA has a certain sensitivity and specificity. The treatment effect of the DA is included in the model through its sensitivity and false positive rate (FPR; same as *1 – specificity*).

The sensitivity corresponds to the probability of correctly predicting a hospitalisation when that would be the next event to be processed in the model. A hospitalisation is avoided in the simulation when it is correctly detected and clinically avoidable, where the latter is approximated by the average for potentially preventable hospitalisations in HF reported in the literature, which is 50% (208). Thus, assuming the sensitivity of the alarm is 0.52 and that 50% of the hospitalisations are clinically avoidable, then 0.52 x 50% = 26% would be the overall probability of avoiding a hospitalisation.

FPR represents the proportion of false positive alarms. Hence, if the FPR of the DA (with daily alarms) were 0.03 and there were 100 days between the previous and the current events simulated in the model, there would be 3 false positive alarms during the period between both events. The false positive alarms are included in the model through the cost of managing those alarms and they are assumed to have no consequences for health outcomes.

In our study we used the DA developed using a multi-resolution analysis signals for diastolic blood pressure and weight collected daily by a non-invasive HTM for predicting hospitalisation published elsewhere (244). The sensitivity and FPR in the base-case analysis were set to the figures reported in that study: 0.52 and 0.03, respectively.

Outpatient visit costs

The office visits reported in the TEN-HMS trial discriminated between general practitioner (GP), nurse, and specialist visits for both UC and HTM (126). We assumed that this partition is representative of the Dutch clinical practice for the population under analysis. Through calculating the weighted average between the product of the type of visit and its reference price in the Dutch costing manual (211), we estimated the costs of an outpatient visit to be €44.50 for UC and €43.30 for HTM (Table 1 in the Online Appendix 6.1).

Costs of other HF-related care provider contacts

The number and type of health care resources used (emergency room visits, office visits, home visits, and telephone calls) during the TEN-HMS trial were reported for UC and HTM for 240 follow-up days (126). TEN-HMS data were also assumed to represent Dutch clinical practice for the HF management. For estimating the costs of

other HF-related care provider contacts we excluded the count on office visits, as these were used separately for estimating the cost per outpatient visit (see above). We converted the resources used during the follow-up period in the TEN-HMS trial (240 days) to yearly rates per patient and we multiplied these figures by the cost for the resource included in the Dutch costing manual (211). The estimated costs of other HF-related care provider contacts per year were €188.38 for UC and €623.61 for HTM (Table 2 in the Online Appendix 6.1).

Hospitalisation costs

The average hospital stay in days in the Netherlands for HF is 8.6 for men and 8.4 for women (233). The gender partition of the population included in the TEN-HMS trial was 78% men and 22% women (126). Using the average cost of a hospital day from the Dutch costing manual (211) and the weighted average of hospital days according to gender, we estimated the average costs per hospitalisation at €4,404.46 (Table 3 in the Online Appendix 6.1).

HTM costs

We used the mid-point of the telemonitoring costs from the range of yearly equipment and service fee and the instalment fee (every 5 years) reported elsewhere (205) for obtaining an yearly cost estimate of €1,257.75 for HTM. Additionally, we used the cost for a GP teleconsultation reported in Dutch costing manual (211) (€18.38) for the cost of managing false positive alarms raised by the DA (Table 4 in the Online Appendix 6.1).

Drug costs

The TEN-HMS database contained information about the drugs used by each individual patient. Every drug that was reported to have been used in more than 5% of the total patients was included in the cost analysis. The daily dose assumptions for each drug were obtained from the figures reported elsewhere (247) and confirmed by expert opinion. The representativeness of the TEN-HMS trial for the Dutch clinical practice for the considered population was discussed above and assumed for the drug use.

The daily drug costs were based on the cheapest option available in the Z-index (245) and calculated applying the following formula from the Dutch costing manual (211): *Drug costs = pharmacists purchase price (Z-index) – clawback (8.3%) + VAT (6%) + pharmacy dispensing fee*. The pharmacy dispensing fee was included by dividing the total fee by the number of units in the considered presentation and multiplying it by the number of units taken daily. The costs for insulin therapy were not available in the Z-index database and were extracted from the literature (248).

The total average drug costs per patient per year were estimated at €286.44 (Table 5 in the Online Appendix 6.1 shows the breakdown of drug costs included in the model).

Informal care costs

The TEN-HMS database contained information on the burden to others reported at baseline for 98.6% of patients. Possible answers were: "no", "very little", "a little", "some", "a lot", and "very much". These were modelled to correspond to 0%, 2%, 4%, 6%, 8%, and 10% of time spend on informal care during a 16 hour day, respectively. After analysing these data by NYHA class we determined that there were no significant differences between classes (Table 6 in the Online Appendix 6.1) and we used the average of the whole population for obtaining informal care costs. The total average cost of informal care per patient per year (€2,098.28) was obtained by multiplying the average hours of informal care per 16-hour day by 365.25 days and by the hourly cost of informal care from the Dutch costing manual (211) (Table 7 in the Online Appendix 6.1).

Travelling expenses

Travelling expenses were calculated based on Kanters et al. (246) and added to the costs of outpatient visits and hospitalisations (Table 8 in the Online Appendix 6.1).

Costs related to productivity losses

Because we used a patient-level simulation model, we were able to include age and gender specific productivity costs for each individual patient until 65 years of age, after which we assumed that patients did not incur further productivity costs.

Productivity losses were assigned to hospitalisations for patients who were considered to be working at baseline. We assumed that a hospitalised patient incurs productivity costs for one whole month, as it seems unlikely that the patient is able to return to work immediately after being hospitalised. We further assumed that the working status does not change during the model, which led to excluding long-term productivity costs from the model. We used the proportion of patients assumed to be working per NYHA class from expert opinion reported elsewhere (247). This working probability of each patient was adjusted using an age-gender-specific net-labour participation rate for the general population (243). The total cost per day was calculated using age-gender-specific data on working hours per week and hourly labour cost (243) (in the Online Appendix 6.1, Table 9 shows the inputs for the calculation of productivity costs and Table 10 an example of the costs incurred by a hypothetical patient).

Future unrelated medical costs

Dutch guidelines require the inclusion of additional costs from unrelated diseases during the life years gained with interventions extending life-expectancy (210). We extracted the estimates of per capita health care expenditures by age and gender from the PAID 3.0 tool and we included those costs for each patient individually during the simulation (242, 249) (Table 11 in the Online Appendix 6.1).

Health outcomes and utilities

QALYs were obtained through weighing life years with patient utility over time. Utilities were attributed to each patient at the start of the simulation according to their NYHA class at baseline and to NYHA class specific utility values reported elsewhere (205) (Table 12 in the Online Appendix 6.1). Utilities change over time with events occurring in the simulation. It was assumed that there were no utility changes resulting from outpatient visits and that hospitalisations resulted in a decrease in utility by a factor of 0.82, following the change in utility observed between NYHA classes reported in another study published for a similar HF population (207). We assumed that the disutility factor from hospitalisations should be limited to three events.

Based on the equation estimated by Ara and Brazier for the utilities for the general UK population (Equation 1 in the Online Appendix 6.1), age-gender-specific utilities attributed at baseline were capped and a decrement factor for ageing was implemented (250).

Cost-effectiveness

Average outcomes per patient were presented for each intervention. The incremental cost-effectiveness ratio (ICER) was calculated as the difference in the average total costs per patient divided by the difference in the average number of QALYs per patient (€/QALY).

Sensitivity and scenario analyses

Parameter uncertainty was assessed via deterministic sensitivity analyses (251). Joint parameter uncertainty was explored through probabilistic sensitivity analysis (PSA), including the parameter-distributions specified in Table 6.2 (252, 253). Following the methodology for addressing uncertainty in DES models published elsewhere (254), PSA was implemented as a double loop: an inner loop, in which a pre-determined number of patients are sampled with replacement from the baseline population, and an outer loop, in which values of the input parameters of the model are randomly drawn. The results from a PSA with an inner loop of 100 patients and an outer loop of

500 iterations were plotted on the cost-effectiveness plane (253, 255, 256). Cost-effectiveness acceptability curves were drawn (92, 257).

Scenario analyses in which key structural assumptions regarding time-to-death and time-to-hospitalisation parametric survival models, time-to-outpatient visits, utilities, and costs were varied to estimate the impact of those assumptions on the outcomes were performed.

Value of information (VOI) analysis

The guidelines for economic evaluations in the Netherlands require the calculation of the expected value of perfect information (EVPI) when the probability that the intervention is cost-effective at the appropriate cost-effectiveness threshold is lower than 100% (210). EVPI per patient is calculated from the average of the maximum net-benefits in each PSA iteration minus the maximum average net-benefit for the interventions considered in the analysis (94, 258, 259). The population EVPI is achieved through multiplying EVPI per patient by the size of the potential population benefiting from the new intervention across the time span for which the recommendation resulting from the VOI analysis is applicable. We assumed five years for the expected applicability of the recommendation and we estimated number of patients eligible for the HTM-based interventions in the Netherlands in the period 2020-2024: 53,140, 55,009, 56,943, 58,946, and 61,019, respectively (233, 260, 261). We discounted EVPI at 4% per year.

Cost-effectiveness of the DA

In the context of the predictive performance of binary diagnostic tests, an ROC curve is the graph that illustrates the diagnostic ability of a binary classifier system by plotting the sensitivity against FPR (*1 – specificity*) at various threshold settings.

In order to properly assess the cost-effectiveness of the DA when added to the HTM intervention, we ran the model at different points of the ROC curve of the DA other than the base-case scenario, thus inferring at which combinations of sensitivity and specificity the diagnostic algorithm would be the most cost-effective. In other words, this analysis intends to determine the operating point at which the threshold of the DA should be set in order to achieve the best balance between costs and health outcomes for the HTM+DA intervention. The values of sensitivity and FPR were measured using the website http://www.graphreader.com/ (161).

Subgroup analyses

We analysed a wide range of subgroups by varying patient and disease characteristics present in Table 6.1. We created two subgroups based on age – patients < and ≥ 65

years of age – and two based on the ejection fraction – patients with less and those with more than 25%. We further analysed patients belonging to each NYHA class separately, creating four more subgroups. Finally, every dichotomous variable generated two subgroups (characteristic present/not present). In total, we analysed 26 different patient subgroups.

Results

Cost-Effectiveness Threshold

The standardised quality-adjusted life expectancy for the population included in the analysis (≈67 years of age and 78% males) is 14.7 QALYs. The total expected undiscounted QALYs accrued with the current standard of care (UC) in the model are 1.16, which represents that 92.1% of normal quality-adjusted life expectancy is lost due to the disease. In this situation, the appropriate cost-effectiveness threshold using the proportional shortfall approach is €80,000 per QALY.

Base-case analysis

The main results of the base-case analysis are summarised in Table 6.3.

Table 6.3 – Results of the base-case analysis

Average outcomes per patient	**UC**	**HTM***	**HTM+DA**
Intermediate outcomes (events per year)			
Outpatient visits	3.60	6.62	6.63
Hospitalisations	1.70	1.64	1.31
Avoided hospitalisations	–	–	0.45**
Death type			
Death in hospital, %	47.2	64.2	58.5
Death (other), %	52.8	35.8	41.5
Final outcomes (discounted)			
Total costs, €	€46,879	€60,343	€65,008
Total life years	2.18	2.96	3.44
Total QALYs	1.12	1.51	1.78
Incremental cost-effectiveness analysis	**HTM vs. UC**	**HTM+DA vs. HTM**	**HTM+DA vs. UC**
Δ€	€13,465	€4,665	€18,129
ΔQALY	0.39	0.26	0.65
Δ€/ ΔQALY	€34,449*	€17,713	€27,712

Abbreviations: DA, diagnostic algorithm; HTM, home telemonitoring; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; UC, usual care.

* Extendedly dominated by HTM+DA.

** Avoided hospitalisations within the HTM+DA intervention group

UC patients experienced approximately 3 outpatient visits per year less than HTM-based interventions. Conversely, HTM results in a decrease of the yearly rate in hospitalisations when compared to UC (1.64 vs. 1.70). This decrease is even more pronounced when the DA is added to HTM, since 0.45 (95% CI: 0; 2.12) hospitalisations per year are avoided due to the DA.

UC is the intervention with the lowest total discounted costs (€46,879), followed by HTM (€60,343), and HTM+DA (€65,008). Patients are expected to survive on average 2.18 discounted years with UC, 2.96 with HTM, and 3.44 with HTM+DA, corresponding to 1.12, 1.51, and 1.78 discounted QALYs, respectively. The hierarchical analysis of the costs and QALYs of the three interventions shows that HTM is extendedly dominated by HTM+DA, as the ICER of HTM compared with UC (€34,449/QALY) is higher than the ICER of HTM+DA (the next, more effective, alternative) compared with UC (€27,712/QALY).

Sensitivity and scenario analyses

Considering the extended dominance of HTM+DA over HTM, the univariate sensitivity analyses were only performed for the HTM+DA versus UC comparison. The results of the five input parameters with the largest effect on the ICER are presented in the tornado diagram in Figure 6.1. All ICERs remained below the €80,000/QALY threshold.

The PSA outcomes plotted in the cost-effectiveness plane for each pairwise comparison show that the great majority of simulations fall in the northeast quadrant, i.e. interventions have higher costs and accrue more QALYs than their comparators (Figure 6.2). The probabilistic ICER between HTM+DA and UC was similar to the deterministic one found in the base-case analysis: €25,864/QALY (95% CI: €15,527; €54,151). The cost-effectiveness acceptability curves for the three interventions show that UC is the alternative expected to be the most cost-effective at low WTP thresholds, HTM is never the most cost-effective intervention, and HTM+DA becomes the intervention most likely to be cost-effective from €25,864/QALY upwards, reaching a 96.0% probability at the appropriate cost-effectiveness threshold of €80,000/QALY (Figure 6.3).

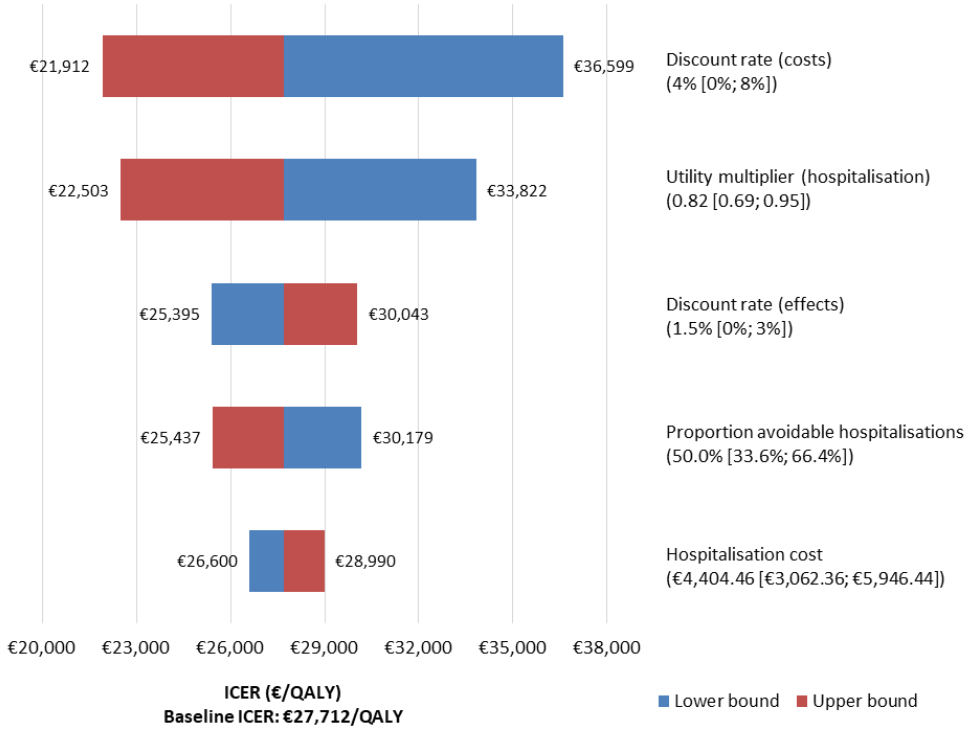


Figure 6.1 – Tornado diagram for the HTM+DA vs. UC comparison

The results of the scenario analyses assessing the structural assumptions of the model are summarised in Table 13 in the Online Appendix 6.1. The scenario with the highest impact on the ICER was the one where a healthcare perspective was taken, which resulted in an ICER between HTM+DA and UC of €14,408/QALY (-48.0% when compared to the base-case analysis). On the opposite direction, the scenario taking all costs from upper bound of 95% confidence intervals was the one with the highest ICER (€31,829/QALY). All ICERs from the scenario analyses stayed below the threshold of €80,000/QALY.

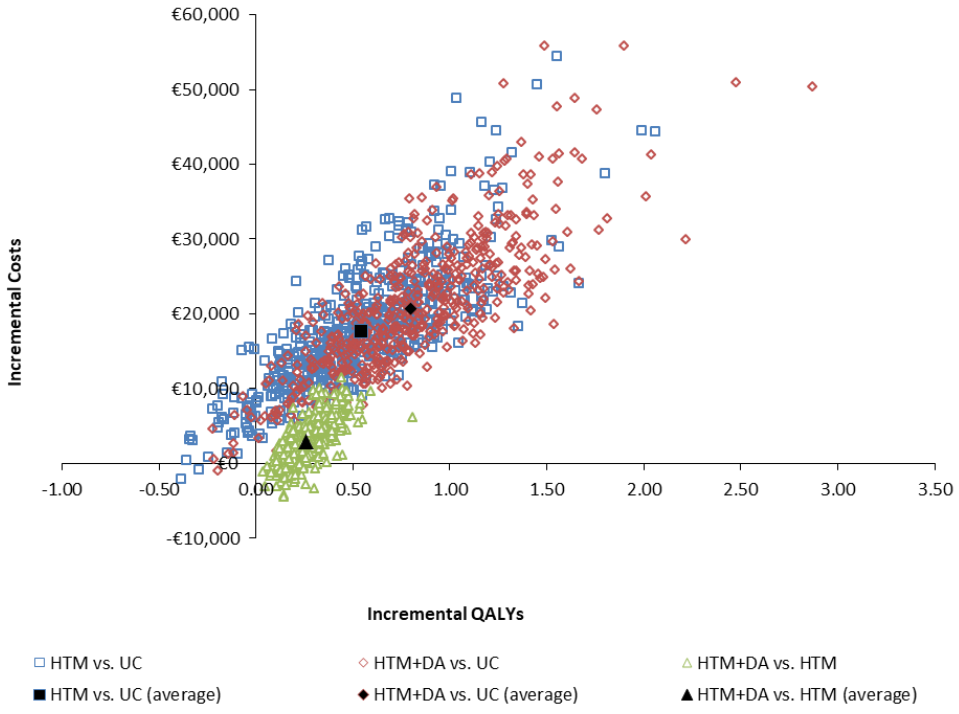


Figure 6.2 – Incremental cost-effectiveness plane

Value of information (VOI) analysis

In the base-case analysis, at the appropriate threshold of €80,000/QALY, the probability of HTM+DA being cost-effective was 96.0%. The calculated EVPI per patient was €341. With an estimated number of patients eligible for the HTM-based interventions in the Netherlands of 253,118 patients (after discounting) for the period 2020-2024, the population EVPI was estimated at €86,383,575.

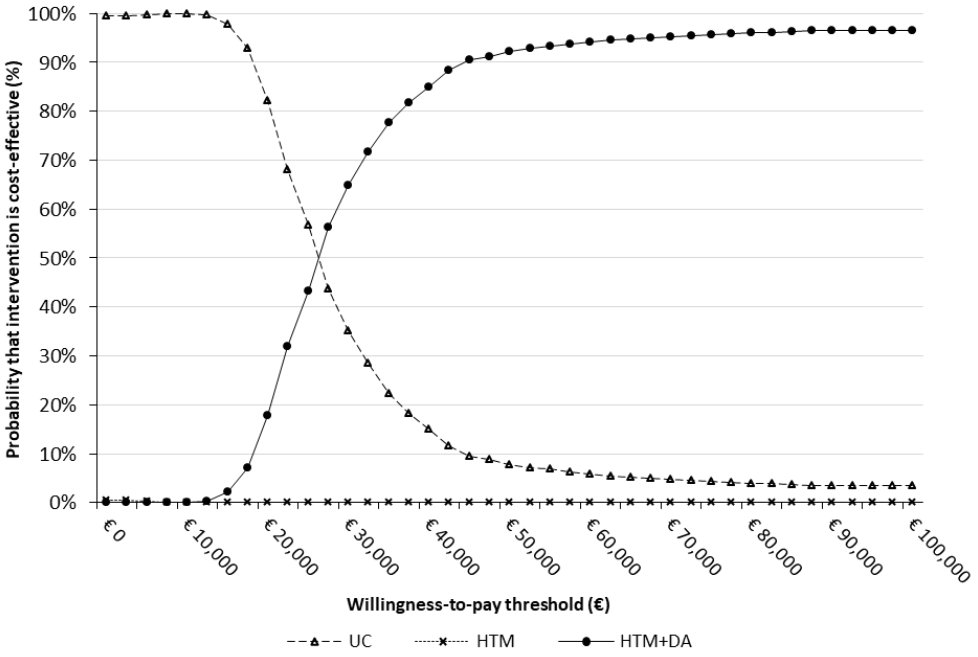


Figure 6.3 – Cost-effectiveness acceptability curves

Cost-effectiveness of the DA

Results for the treatment scenarios assuming different characteristics of the DA are presented for the comparison of HTM+DA with UC in Table 6.4. Increasing the sensitivity of the DA by setting a lower threshold for the alarm to go off, which entails an increase in FPR (decreased specificity), resulted in a higher number of avoided hospitalisations, life years and QALYs, but with higher costs. Alternatively, decreasing the sensitivity (i.e. setting a higher threshold for the alarm) resulted in lower costs, but with worse health outcomes. From the scenarios tested, the most cost-effective was scenario 3, where sensitivity was set to 0.600 and FPR to 0.068. In the scenarios testes, moving away from that point in either direction of the ROC curve resulted in higher ICERs (ICER range: €25,734/QALY; €35,560/QALY).

Table 6.4 – Results of the scenario analyses for the diagnostic algorithm

Average outcomes per patient (HTM+DA)	**DA scenarios**					
	1 Sens: 0.200 FPR: 0.007	**2** Sens: 0.400 FPR: 0.024	**BC** Sens: 0.520 FPR: 0.030	**3** Sens: 0.600 FPR: 0.068	**4** Sens: 0.800 FPR: 0.194	**5** Sens: 0.950 FPR: 0.562
Intermediate outcomes (events per year)						
Outpatient visits	6.63	6.64	6.63	6.62	6.62	6.63
Hospitalisations	1.52	1.36	1.31	1.23	1.10	1.00
Avoided hospitalisations	0.18	0.33	0.45	0.56	0.76	0.92
Final outcomes						
Total costs, €	€62,085	€63,394	€65,008	€64,163	€71,016	€82,108
Total life years	3.14	3.29	3.44	3.43	3.73	3.99
Total QALYs	1.61	1.70	1.78	1.80	1.96	2.11
ICER						
vs. UC, €/QALY	€30,984	€28,881	€27,712	€25,734	€29,004	€35,560
change vs. base-case	+11.8%	+4.2%	0.0%	-7.1%	+4.7%	+28.3%

Abbreviations: BC, base-case; DA, diagnostic algorithm; FPR, false positive rate; HTM, home telemonitoring; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; Sens, sensitivity; UC, usual care.

Subgroup analyses

The summary of cost-effectiveness results for the subgroup analyses is presented in Table 6.5. Because each subgroup is created from a subset of the population in the TEN-HMS database (126), the characteristics of the baseline population for each subgroup may differ. The baseline patient and disease characteristics of the model population for each of the analysed subgroups are presented in Tables 1 to 26 in the Online Appendix 6.2. All the ICER changes versus the base-case concern the comparison between HTM+DA and UC.

The subgroup of NYHA class IV patients registered the highest deviation from the base-case analysis results, with an ICER of €52,727/QALY (+90.3%). On the contrary, the subgroups with the better cost-effectiveness ratios were patients younger than 65 years of age and patients belonging to NYHA class I (€22,830/QALY [-17.6%] and €22,870/QALY [-17.5%], respectively).

Although many other subgroups did not show such a high variation in the ICER – as this is a ratio that depends on the simultaneous variation of costs and QALYs for each of the interventions being compared –, large differences in final outcomes were observed for some subgroups. Males (especially when compared to females), patients from NYHA class III, with diabetes, with chronic obstructive pulmonary disease, not on beta-blocker medication, not on angiotensin-converting enzyme medication, with history of myocardial infarction, and with history of chronic atrial fibrillation recorded a considerable decrease in QALYs for both HTM+DA and UC. For those subgroups, given that we are dealing with dichotomous variables, the complementary subgroups resulted in higher QALYs (i.e. better health outcomes), with the exception of smokers vs. non-smokers, where the comparison between those two subgroups showed small differences in QALYs and costs.

For all the subgroups showing a decrease in QALYs, a decrease in costs was also observed. This fact corroborates the positive correlation between costs and effects that was noticeable in the incremental cost-effectiveness plane show in Figure 6.2. Hence, a decrease in life expectancy and, therefore, QALYs, is associated with increased ICERs when compared to the base-case analysis.

Table 6.5 – Subgroup analyses: summary of cost-effectiveness results

	Subgroup*	Costs (€)			QALYs			ICER (€/QALY)	
		UC	HTM**	HTM+DA	UC	HTM**	HTM+DA	HTM+DA vs. UC	% vs. base-case
-	Baseline population	€46,879	€60,343	€65,008	1.12	1.51	1.78	€27,712	0.0%
1	Age < 65 years old	€59,543	€75,311	€79,144	1.78	2.25	2.64	€22,830	-17.6%
2	Age ≥ 65 years old	€39,380	€52,035	€56,483	0.82	1.14	1.32	€34,368	+24.0%
3	Ejection fraction < 25%	€45,516	€60,745	€64,906	1.22	1.67	1.94	€26,813	-3.2%
4	Ejection fraction ≥ 25%	€46,843	€61,279	€65,606	1.06	1.39	1.65	€31,372	+13.2%
5	NYHA class I	€53,679	€72,656	€77,377	1.84	2.51	2.88	€22,870	-17.5%
6	NYHA class II	€48,659	€64,094	€67,515	1.24	1.63	1.90	€28,827	+4.0%
7	NYHA class III	€43,142	€51,046	€54,454	0.80	1.00	1.18	€29,759	+7.4%
8	NYHA class IV	€36,821	€45,218	€48,957	0.38	0.50	0.61	€52,727	+90.3%
9	Gender: male	€45,762	€57,518	€61,122	1.08	1.36	1.60	€29,777	+7.5%
10	Gender: female	€51,148	€68,937	€75,954	1.53	2.05	2.38	€29,038	+4.8%
11	Smoker: yes	€49,819	€62,956	€64,973	1.18	1.48	1.73	€27,765	+0.2%
12	Smoker: no	€45,741	€60,614	€64,392	1.13	1.49	1.74	€30,208	+9.0%
13	Diabetes: yes	€43,213	€55,144	€59,211	0.96	1.26	1.48	€30,624	+10.5%
14	Diabetes: no	€48,611	€60,287	€65,193	1.27	1.59	1.86	€27,980	+1.0%
15	COPD: yes	€39,386	€47,599	€52,293	0.80	1.02	1.24	€29,560	+6.7%
16	COPD: no	€49,180	€67,014	€70,128	1.23	1.67	1.92	€30,105	+8.6%
17	Recent diagnosis: yes	€54,103	€69,207	€74,122	1.53	1.90	2.20	€29,748	+7.3%

18	Recent diagnosis: no	€42,619	€53,123	€56,272	0.91	1.18	1.39	€28,567	+3.1%
19	No beta-blocker medication: yes	€38,967	€48,709	€50,661	0.70	0.92	1.09	€29,830	+7.6%
20	No beta-blocker medication: no	€51,211	€67,252	€71,213	1.41	1.85	2.15	€27,127	-2.1%
21	No ACE inhibitor medication: yes	€39,967	€52,165	€54,888	0.76	1.04	1.21	€32,921	+18.8%
22	No ACE inhibitor medication: no	€47,208	€61,294	€65,897	1.20	1.57	1.84	€29,424	+6.2%
23	Myocardial infarction: yes	€43,366	€57,360	€61,261	0.99	1.33	1.56	€30,958	+11.7%
24	Myocardial infarction: no	€51,222	€64,252	€69,338	1.41	1.79	2.10	€26,195	-5.5%
25	Chronic atrial fibrillation: yes	€38,205	€49,469	€53,452	0.72	1.02	1.19	€32,415	+17.0%
26	Chronic atrial fibrillation: no	€50,164	€64,086	€68,856	1.31	1.71	2.01	€26,812	-3.2%

Abbreviations: ACE, angiotensin-converting enzyme; BMI, body mass index; COPD, chronic obstructive pulmonary disease; DA, diagnostic algorithm; HTM, home telemonitoring; ICER, incremental cost-effectiveness ratio; NYHA, New York Heart Association; QALY, quality-adjusted life year; UC, usual care.

* Because each subgroup is created from a subset of the population in the TEN-HMS database (126), the characteristics of the baseline population for each subgroup may differ. The baseline patient and disease characteristics of the model population for each of the analysed subgroups are presented in Tables 1-26 in the Online Appendix 6.2.

** HTM is extendedly dominated by HTM+DA in all analysed subgroups. The ICER comparison against the base-case is only shown for HTM+DA vs. UC.

Discussion

This study aimed at assessing the cost-effectiveness of HTM and a DA in the management of heart failure in the Netherlands and it used a previously validated patient-level discrete event simulation model (238) for analysing three separate interventions: UC, HTM, and HTM+DA. The base-case analysis determined that HTM is extendedly dominated by HTM+DA, with the latter intervention being cost-effective versus UC at a deterministic ICER of €27,712 per QALY gained (Table 6.3).

The cost-effectiveness of the DA was more carefully examined through creating various scenarios with different values for sensitivity and FPR from the ROC curve published by Koulaouzidis et al. 2016 (244). Those scenarios generated model outcomes that allowed for comparing the ICER of HTM+DA vs. UC at various thresholds of the DA (Table 6.4), thereby assessing the inherent trade-off between false positives and false negatives in cost-effectiveness terms. In the particular case of this study, false positives correspond to alarms that were incorrectly raised, as the patient would not have been hospitalised, while false negatives represent alarms that were incorrectly not raised and thus did not possibilitate avoiding a hospitalisation. In the DA scenarios tested, scenario 5 minimises false negatives at the expense of increasing false positives: in this case, more hospitalisations are avoided and QALYs increase as a consequence, but at a higher cost, as there are more false positives, which entail costs without any additional health benefits. Conversely, scenario 1 minimises false positives at the expense of increasing false negatives: in this situation, less hospitalisations are avoided – due a higher number of false negatives –, leading to fewer QALYs and fewer costs, as the false positives drop. Although both false positives and false negatives are undesirable, there is an optimal point in terms of cost-effectiveness that represents the balance between sensitivity and FPR within the ROC curve in terms of generated QALYs and associated costs. In our analysis, scenario 3 is the one closer to this optimal point, as it leads to the lowest ICER of HTM+DA against UC. Because sensitivity and specificity are not independent variables, the optimal point must be found iteratively, as there is a great deal of calculations running in the background of the model and it would be mathematically very hard to solve all the equations for those variables.

The subgroup analyses showed considerable variation in the ICERs of HTM+DA vs. UC (Table 6.5), with the highest ratios being recorded for the subgroups of patients ≥ 65 years of age and those in NYHA class IV. A large variation in costs and QALYs was also observed, even when the resulting ICER did not change much from the base-case analysis for the HTM+DA vs. UC comparison, which may be attributed to the positive correlation between costs and effects observed in the subgroup analyses. It was also observed that complementary subgroups (with the exception of smokers/non-

smokers) went in opposite directions in relation to final outcomes (e.g., lower QALYs and costs for patients with history of myocardial infarction contrasted with higher QALYs and costs for patients without any history of myocardial infarction). Concerning the subgroup analyses, it is critical to emphasise that the interpretation of subgroups in patient-level simulation models is a sensitive matter, as every subgroup created from the baseline population – by restricting the variables of interest to values compatible with the subgroup being analysed – is likely to have different patient and disease characteristics when compared to the model population used in the base-case analysis (Tables 1 to 27 in the Online Appendix 6.2). For instance, NYHA class IV patients are also older on average than the baseline model population (Table 14 versus Table 27 in the Online Appendix 6.2). Hence, the outcomes from the model and their variation from the base-case analysis in that situation are not only depending on the impact of NYHA IV, but also of all other patient and disease characteristics that change in the subgroup population when compared to the base-case population. Thus, the correct interpretation of subgroup analyses requires a link with the patient and disease characteristics than can be correlated with the particular characteristic changing in any given subgroup.

The cost-effectiveness analysis presented in this paper relied on several distinguishing features of the Dutch economic evaluation guidelines: the adoption of a societal perspective, the calculation of productivity losses using the friction cost-method, differential discounting, the inclusion of caregiver burden on the cost side of the economic evaluation, the incorporation of indirect medical costs of life-years gained, and a VOI analysis. Deterministic sensitivity analyses and scenario analyses showed that the model results were robust to the variation of most parameters (Figure 6.1) and to most changes in structural assumptions, with the highest change in the ICER resulting from taking a healthcare perspective in the analysis. This worsening of the ICER is due to the increased costs that are not matched by an increase in life-expectancy and QALYs. PSA results revealed a 96.0% chance that HTM+DA being cost-effective at the appropriate threshold of €80,000/QALY (as determined by the proportional shortfall method). At that threshold, the calculated EVPI per patient, i.e. the monetary consequence of making a wrong decision, was €341. Given the high estimated number of eligible patients for the HTM-based interventions in the Netherlands for the period 2020-2024, the monetary consequence of making a wrong decision at a population level was estimated at €86,383,575.

Comparison to other studies

To our knowledge, this is the first study using a health economic patient-level simulation model for assessing the cost-effectiveness of a HF intervention in the

Netherlands. Concerning the intervention, two studies have also assessed the cost-effectiveness of HTM in the Netherlands: Boyne et al. (2013) (262) and Grustam et al. (2018) (205). Boyne et al. performed a trial-based economic evaluation of the Telemonitoring in Heart Failure (TEHAF) study (263, 264), a prospective open label, multicentre, randomised controlled trial with blinded endpoint evaluation, conducted at three hospitals in The Netherlands. The results of this study can hardly be compared to ours. Firstly, because the population in the TEHAF-study was in better health state than the one in the TEN-HMS study (e.g., mean ejection fraction of 36% vs. 25%), and secondly, because the time horizon of their study was only one year, which cannot properly capture the lifetime change in costs and effects between the interventions since patients are expected to survive more than one year. Grustam et al. used a Markov cohort model with most of the data coming from the TEN-HMS study for assessing the cost-effectiveness of HTM compared to UC. They took a third-party payer's perspective and direct comparison of results with that study would be unwise and uninformative. However, in the scenario analysis where we took a healthcare perspective (scenario 23 in Table 13 in the Online Appendix 6.1), we estimated similar costs: €16,034 for UC and €25,433 for HTM vs. €14,414 and €27,186, respectively, found by Grustam et al. (205). The ICERs, however, were different, because we estimated fewer QALYs in our study. One possible explanation is the assumption by Grustam et al. that the transition probabilities measured in the time frame of 240 to 450 days in the TEN-HMS study continue unaltered for 20 years, which, given the mean age of 67 years of the patients included in the model and their very poor health state, seems unlikely. This assumption may have overestimated survival in their study. Another possible explanation for the aforementioned difference is the potential underestimation of survival in our study due to the regression equation for dying in hospital. The regression equation calculating time-to-death predicts all-cause mortality. Thus, patients dying in hospital may result in some type of double counting of mortality due to the inherent imprecision of data-driven estimates. If predictions were 100% accurate, the model would predict time of death flawlessly – which never happens in practice. However, given the higher number of hospitalisations experienced by patients in the intervention arms – due to their increased survival –, the cost-effectiveness estimates, if anything, are conservative.

The findings in our study of lower mortality and hospitalisations with HTM-based intervention when compared to UC are consistent with the results previously published in two network-meta analyses (112, 265). Regarding costs, we found an increase in total costs with HTM when compared to UC. In the review by Inglis et al. the authors identified three studies reporting costs for HTM vs. UC; one reported a decrease in costs and two reported increases in cost, due both to the cost of the intervention and to increased medical management (265).

It is worthwhile mentioning that the structure of the model used in our study allowed exploring the impact of adding a DA to the HTM intervention. This is a critical aspect of our study, as it is the first to assess the cost-effectiveness of a DA in the context of chronic disease management. Although we have used the concept in the context of a HF intervention, it could be adapted for other disease areas. This subject has been discussed in the publication regarding the validation of the model used in the present study (238).

Limitations

The first limitation stems from the TEN-HMS study dating from 2005, which results in a large enough period for medical practice to have changed, especially since we are talking about technologies which are developed at a fast pace. The experience that results from the continuous use of these technologies can ultimately have an impact on their effectiveness and cost-effectiveness. Still related to the TEN-HMS study, drug use and their costs report to standards in the time of the trial. Even if standards in terms of therapeutic classes are not necessarily different, the drugs used are older and they are likely cheaper than the more recent alternatives as of today (this impact was assessed via scenario analysis). And finally, there could be some variation in healthcare systems between patients included in the TEN-HMS study (UK, the Netherlands, and Germany), which is not accounted for in the model.

The second limitation relates to the DA ROC curve used for the analysis. Since the ROC curve was not obtained using the same population or HTM system, we assumed that the different levels of diagnostic accuracy of the DA, i.e. the different points of the ROC curve, would be applicable to the population in our model as well. The population used in the study by Koulaouzidis et al. (244) seemed to be in a better health state than the one in the TEN-HMS study (126) (e.g., ejection fraction of 36.6% vs. 25.1%). Ideally, we would have a DA constructed with the TEN-HMS data, as we would want to optimise the threshold of a DA that would have been designed with the same HTM system. In that way, we could use the data generated by that system for continuously improving predictions of hospitalisation and, consequently, improving the cost-effectiveness of the HTM+DA intervention.

Recommendations for future research

The model could include individual drug costs and optimise the medication used at each processed event. For that to happen, it should update patient characteristics at those events in order to define the correct medication for each patient. In doing so, the model would also capture the drug costs more accurately.

Further research must be done in order to better describe DAs and the follow-up actions they entail in the clinical practice and disease pathways. While the DES

framework allowed for the assessment of the cost-effectiveness of the DA, the potential of those models opens enormous possibilities for designing a model with highly detailed disease pathways for clinicians and decision makers who are less familiar with decision modelling in the context of the economic evaluation of health technologies. However, the increased complexity of models would come at the expense of the need for patient-level data for building and validating the model. Theoretically, all the patient pathways after an alarm could be included in a DES framework. The question would be whether there would be reliable data on the outcomes for each of the pathways that could be conceived for reacting to an alarm. As it is widely described in the health economic literature, models should abide by the principle of parsimony, i.e. as simple as possible in order to accurately reflect the problem under analysis and allowing for making an informed decision.

Conclusions

Although increased costs of adopting HTM and DA in the management of HF may seemingly be an additional strain on scarce health care resources, the results of this study demonstrate that, by increasing patient life expectancy by 1.28 years and reducing their hospitalisation rate by 23% when compared to UC, the use of these technologies may be seen as an investment, as HTM+DA extendedly dominates HTM and generates an extra QALY for a €27,712 investment. At the appropriate cost-effectiveness threshold of €80,000/QALY resulting from the proportional shortfall methodology used in the Dutch economic evaluation guidelines, HTM+DA has a 96.0% probability of being cost-effective.

Appendices

Online Appendix 6.1

Available at:

https://drive.google.com/file/d/14wGfQN7cdnx5IPF1iAVMYq_9Zmr2I8Zr/view?usp=sharing

Online Appendix 6.2

Available at:

https://drive.google.com/file/d/1kvSsLgLHi2FsIVRsqzME3xICp9Dlg_V-/view?usp=sharing

Chapter 7

Medical devices rise: “We demand the same rights!”

Fernando Albuquerque de Almeida and Mariana Ricardo

Health Policy and Technology (Submitted)

Abstract

Objectives: Understanding whether the regulatory framework creates different standards for medical devices and drugs, assessing whether there is evidence on the impact of these standards on clinical and HTA research, and reflecting on the findings in order to propose legislative changes that could promote an integrated evidence-based assessment system that could arguably result in a more efficient allocation of resources in the healthcare systems.

Methods: This study reviewed and compared the legal framework for the approval of medical devices and drugs in the EU, with a particular focus on the changes brought by Regulation (EU) 2017/745. It further investigated the available information on manufacturer sponsored clinical studies and HTA-supported recommendations for medical devices and drugs.

Results: The review of the legislation identified different standards for approval of devices and drugs on their quality, safety, and performance/efficacy dimensions. We found substantially lower number of manufacturer sponsored clinical studies and HTA-supported recommendations for medical devices versus drugs. Further, there is an indication of lower standards of evidence used in recommendations for medical devices.

Conclusions: Policy changes ought to be implemented in order to promote an integrated evidence-based assessment system for a better allocation of resources in healthcare, namely: a consensual classification of medical devices from an HTA perspective, which could be used as a guide for generating outcomes in clinical investigation, and the adoption of conditional coverage practices including mandatory post-approval evidence development for performing periodic technology assessments.

Introduction

A medical device (MD) can be defined as any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, intended for use in the diagnosis of disease or other conditions, in the cure, mitigation, treatment, or prevention of disease, or in the investigation, replacement or modification of the anatomy or of a physiological process, and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means (266). In the European Union (EU), MDs are categorised in classes, based on their intended purposes and inherent risks, which determine the regulatory controls necessary to provide a reasonable assurance of safety and performance.

Attention given to MDs has been increasing steadily, with the number of patent applications for medical technology to the European Patent Office (EPO) growing 32.5% between 2010 and 2019 (267). MDs represent a significant share of the healthcare expenditures in the EU: in 2018 the total spending with MDs and in vitro diagnostics was estimated at around 120€ billion, or 7.4% of the money spent on healthcare (if we exclude in-vitro diagnostics, 108.6€ billion, or 6.7%), with the EU5 countries – France, Germany, Italy, Spain, and the United Kingdom (UK)[1] – representing 70% of the total expense with MDs in the EU (268).

Bringing a device to market takes on average 3 to 7 years, compared with an average of 12 years for drugs (269). Evidence shows that the EU faces challenges ensuring that only safe and effective devices reach the market and monitoring their real-world utilization, and there is a lack of quantitative evidence assessing MD regulation (270, 271).

Traditionally MDs have been less regulated than drugs and the standards of evidence collection for placing them on the market are generally lower, which may ultimately hinder comparison between these health technologies, especially when they have the same goal and are targeted at the same disease and/or population (272-274). Health technology assessment (HTA), aimed at informing decisions on the adoption of new health technologies, broadly used for drugs, usually comes across methodological issues due to scarce evidence on clinical effectiveness when assessing MDs, as the regulation for their access to the market focuses on safety and performance (275).

[1] Although the UK has formally withdrawn from the EU as of 1 February 2020 at 00:00, for the purpose of this study the UK will be considered a EU5 country, as most of the data included in the study pertain to a period in which the UK was a Member State of the EU.

This study reviews and compares the legal framework in the EU for the approval of MDs and drugs, with a particular focus on the changes brought by new Regulation (EU) 2017/745, which aims to ensure the smooth functioning of the internal market as regards MDs, taking as a base a high level of protection of health for patients and users, and taking into account the small and medium-sized enterprises that are active in this sector, at the same time it sets high standards of quality and safety for medical devices in order to meet common safety concerns as regards such products. Additionally, this study seeks to compare the available information on the clinical research and the HTA-based recommendations for the health technologies under analysis. The main goal of the study was to understand whether the regulatory framework creates different standards for MDs and drugs, to assess whether there is evidence on the impact of these standards on clinical and HTA research and to critically reflect on the findings in order to propose legislative changes that could promote an integrated evidence-based assessment system that would arguably result in a more efficient allocation of resources in the healthcare systems, as both MDs and drugs are usually covered by the same limited healthcare budget.

Methods

EU Law definitions

EU Law is divided into primary and secondary legislation (276). Primary legislation is constituted by the treaties, which are binding agreements between the member States to the EU and contain the ground rules for EU action (Treaty on the European Union and Treaty on the Functioning of the European Union, Treaty establishing the European Atomic Energy Community, and the Charter of Fundamental Rights of the European Union). Secondary legislation includes binding legal instruments such as regulations, directives, and decisions. Additionally, there is soft law which corresponds to non-binding instruments: resolutions, opinions, and other instruments such as EU institutions' internal regulations (277). More specifically on the binding secondary legislation, regulations are legislative acts that must be applied in their entirety across the EU and that become automatically binding on the date they take effect, whereas directives are legislative acts that set out goals/results that addressed Member States must achieve, although they may give leeway to each country for devising its own laws for reaching those goals/results (278).

Identification and review of EU Law for approving medical devices and drugs

For MDs, in order to assess the EU legal framework for placing them on the market – the term equivalent to obtaining a marketing authorisation (MA) in the drug realm –, we extracted regulations and directives from EUR-Lex, the online gateway to EU Law (276).

For drugs, we analysed Directive 2001/83/EC on medicinal products for human use, as amended (279), and Regulation (EC) 726/2004 laying down EU procedures for the authorisation of medicinal products for human use, as amended (280), since they constitute the cornerstone for drug approval in the EU. Both MAs granted following centralised procedures, pursuant to Regulation (EC) 726/2004, and MAs granted through decentralised and mutual recognition procedures, in accordance to Directive 2001/83/EC, are valid throughout the EU and confer the same rights and obligations in all Member States.

The existence of secondary EU legislation containing the definition of the general requirements on quality, safety, and performance/efficacy (performance is the term preferred for MDs, while efficacy is the term used for drugs) for ensuring access to market means that either through the transposition of the rules in the directives or the direct application of regulations the same standards apply to all Member States.

Clinical research and HTA-supported recommendations for medical devices and drugs

In order to assess how clinical research for MDs compares with that for drugs, we searched the ClinicalTrials.gov database for the clinical studies registered for both MDs and drugs as the intervention type (281). We searched for phase II and phase III clinical trials (interventional studies) funded by the industry, as we assumed that these are the studies leading to or containing the most relevant data for submissions to regulatory approval of drugs (and MDs) and/or to HTA submissions for pricing and reimbursement (P&R) purposes. We reported the number of clinical trials registered in the ClinicalTrials.gov database by type of intervention, chronologically by date of study start (before 2010 and yearly after that). Although we intended to perform a similar analysis on the EU Clinical Trials Register (266), this database only contains information on interventional clinical trials on drugs conducted in the EU or the European Economic Area.

In regards to the comparative analysis of the HTA-based recommendations by jurisdiction for MDs and drugs, we firstly identified the relevant HTA body for each of the EU5 countries: for France, the Haute Autorité de Santé (HAS); for Germany, the Gemeinsamer Bundesausschuss (G-BA) and the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); for Italy, the Agenzia Italiana del Farmaco (AIFA) and the Agenzia Nazionale per i Servizi Sanitari Regionali (AGENAS); and for the UK, the National Institute for Health and Care Excellence (NICE). In Spain the provision of healthcare services is decentralized and thus the responsibility of the autonomous communities; there is no central HTA agency. We used the Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) – a collaboration network between the HTA agencies from the autonomous communities, with a common methodology and under the principle of mutual recognition and cooperation – as the data source for our analysis on medical devices. Additionally, we searched the website of the Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) – the central drug regulating agency belonging to the Spanish Ministry of Health – for extracting their therapeutic positioning reports on drugs.(282)

The websites of the previously mentioned institutions were searched for all publicly available assessments/recommendations (283-290) (the details of the website searches and inclusion/exclusion criteria are provided in the Online Appendix 7.1). The number of results was reported separately for MDs and drugs. The type of the publicly available documentation was indicated, as there may be some inter-jurisdiction variation regarding the assessment or type of recommendation made.

Results

Identification and review of the relevant legislation concerning medical devices

The complete list of the legislative acts retrieved from EUR-Lex is presented in the Online Appendix 7.1. After assessing the title and the scope of the retrieved legislation, we reviewed in detail the following legislative acts, which contain all the relevant information for the description of the regulatory framework in the EU: Council Directive 90/385/EEC of 20 June 1990 relating to active implantable medical devices (291) (AIMDD), Council Directive 93/42/EEC of 14 June 1993 concerning medical devices (292) (MDD), and Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017, on medical devices (293) (MDR).

The MDR was approved in 2017 and the due date of application of the large majority of its provisions was scheduled for 26 May 2020. However, the date of application was postponed to 26 May 2021 by Regulation (EU) 2020/561 and the MDR has repealed both AIMDD and MDD on that date (294). Regulation (EU) 2020/561 simultaneously adapted transitional provisions of Regulation (EU) 2017/745 that would otherwise no longer apply. This postponement is explained by the "unprecedented magnitude of the current challenges [COVID-19 outbreak and the associated public health crisis], and taking into account the complexity of Regulation (EU) 2017/745, it is very likely that Member States, health institutions, economic operators and other relevant parties will not be in a position to ensure the proper implementation and application." Considering that MDs regulated by the AIMDD and the MDD may still be placed on the market, we reviewed those directives in addition to the MDR. Any relevant difference in the scope of the present analysis introduced by the MDR is appropriately pointed out.

Classification of medical devices and definition of the relevant technologies for the comparison with drugs

Considering the purpose of comparing regulatory standards for MDs and drugs, it is relevant to categorise MDs into interventional and non-interventional devices. Drugs or medicinal products are defined as substances or combination of substances presented as having properties for treating or preventing disease in human beings, or any substances or combination of substances which may be used in or administered to human beings either with a view to restoring, correcting or modifying physiological functions by exerting a pharmacological, immunological or metabolic action, or to making a medical diagnosis.

The review of AIMDD, MDD, and MDR in regards to their definition of MDs allowed for establishing that distinction, with interventional MDs being the ones intended for treatment of a disease, injury or disability, whereas non-interventional MDs can be defined as the ones directed at diagnosis, monitoring, prognosis or alleviation of disease, injury or disability. We found this distinction to be further supported by the definition of MDs contained in the AIMDD, and the classification rules contained in Annex IX of the MDD and Annex VIII of the MDR, where the reference to active therapeutic devices is found: any device used, whether alone or in combination with other devices, to support, modify, replace or restore biological functions or structures with a view to treatment or alleviation of an illness, injury or disability. Through the analysis of those annexes, we further ascertained that interventional MDs belong to classes IIa (all active therapeutic devices intended to administer or exchange energy), IIb (active devices intended to emit ionising radiation and intended for therapeutic interventional radiology), or III (implantable devices intended to be used in direct contact with the heart, the central circulatory system or the central nervous system).

The classification of interventional MDs helps to set the boundaries for reviewing the rules for placing them on the market, with only relevant standards for the comparison between MDs and drugs being assessed.

Regulatory framework of medical devices across jurisdictions and its comparison to drugs

In the EU drugs are granted marketing authorisation – Article 6 of Directive 2001/83/EC and Article 3 of Regulation (EC) 726/2004 –, whilst MDs are subject to conformity assessment procedures aimed at demonstrating compliance with the requirements of the AIMDD, the MDD, or the MDR – Article 9 of the AIMDD, Article 11 of the MDD, and Article 52 of the MDR. The applicable conformity assessment procedure is determined in accordance with the device classification, with manufacturers being able to select their preferred route. Conformity assessment procedure routes for MDs of classes IIa and of classes IIb and III are shown in Figure 7.1 and Figure 7.2, respectively.

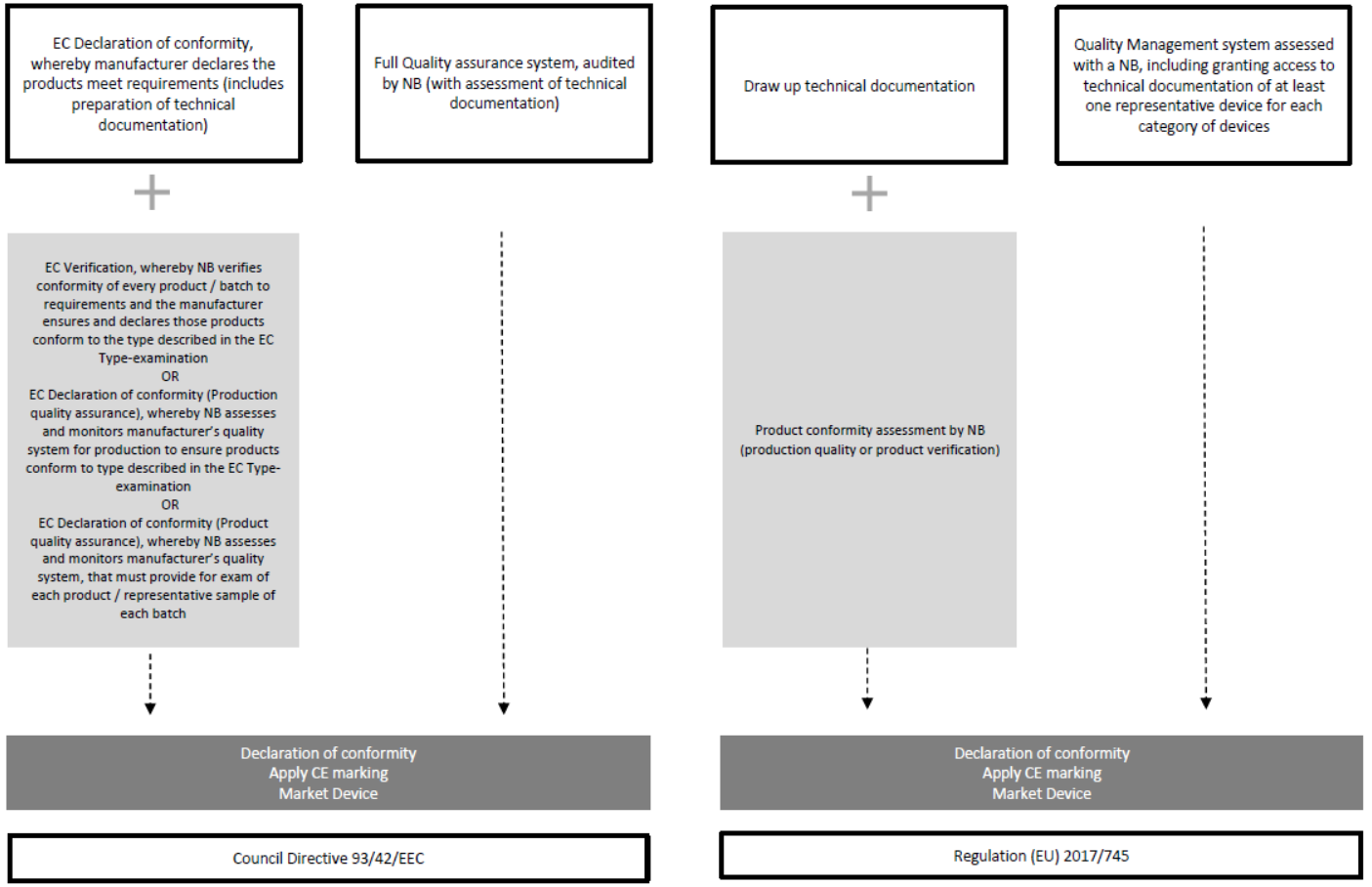


Figure 7.1 – Class IIa medical devices conformity assessment procedure routes
CE, Conformité Européene (European Conformity); EC, European Community; EEC, European Economic Community; EU, European Union; NB, Notified Body.

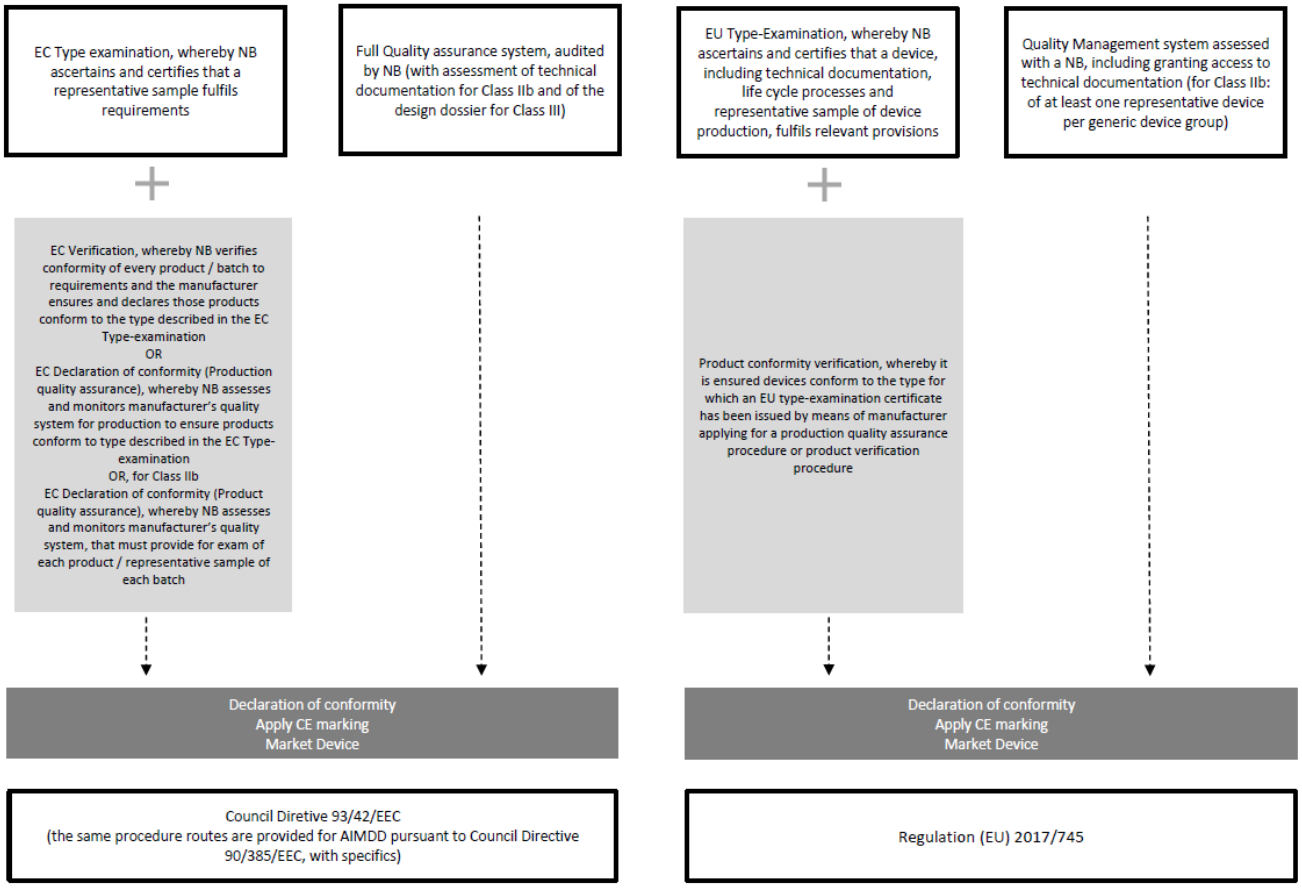


Figure 7.2 – Class IIb + III medical devices conformity assessment procedure routes
AIMDD, Active Implantable Medical Devices Directive; CE, Conformité Européene (European Conformity); EC, European Community; EEC, European Economic Community; EU, European Union; NB, Notified Body.

The relevant body for performing third-party conformity assessment activities is the Notified Body (NB), an independent third party designated by Member States (Article 11 and Annex 8 of the AIMDD, Article 16 and Annex XI of the MDD, and Articles 35-50 and Annex VII of the MDR). MDs found in conformity to EU rules are then registered with the competent authority in each EU country where they are placed on the market. Marketing authorisations for drugs are granted by the European Commission (centralised procedure) or by the medicines regulatory authority in the relevant Member State (mutual recognition and decentralised procedures) (Article 3 of Regulation (EC) 726/2004 and Article 6 of Directive 2001/83/EC, respectively).

If considered to meet applicable performance, safety, design, and manufacture requirements, MDs must bear the CE marking of conformity, which ensures the device may be freely marketed anywhere in the EU without further control (295). Confirmation of conformity with requirements, and the evaluation of the side-effects and of the acceptability of the benefit/risk ratio of MDs must be based on clinical data (AIMDD and MDD) and on clinical data providing sufficient clinical evidence. Under the MDR it is the responsibility of the manufacturers to specify and to justify the level of clinical evidence necessary to demonstrate conformity: to this end, a clinical evaluation shall be planned, conducted, and documented.

We noted both the AIMDD and the MDD required clinical investigations to be performed for class III medical devices and implantable devices (which may fall under classes IIa or IIb), unless it was duly justified to rely on existing clinical data. However, the MDR determines that for the same devices and as a general rule data should be sourced from clinical investigations that have been carried out under the responsibility of a sponsor. This requirement, however, shall not apply to implantable and Class III devices placed on the market in accordance with the MDD.

Regulatory framework of medical devices and drugs in their quality, safety, and performance/efficacy dimensions

We analysed the requirements for the approval of MDs and drugs in their quality, safety, and performance/efficacy dimensions – performance is the term preferred for MDs, while efficacy is the term used for drugs. The repeal of the directives regulating MDs and the adoption of a regulation seems to reflect the EU's consideration that the proper functioning of the MD market could benefit from an extensive regulatory framework review aimed at achieving higher levels of health and safety. The nature of the legislative act adopted to review the EU MD framework is a non-negligible indicator that EU lawmakers took into consideration the growth of the use of these health technologies and their impact on the lives of patients (recital (1) of the MDR). The MDR reinforces key elements of the regulatory approach set out in the repealed

directives, such as clinical investigations and clinical data and the designation and supervision of the NBs performing conformity assessments.

The MDR introduces more stringent requirements regarding clinical investigations, which saw their importance upgraded from a non-exhaustive annex in the directives, containing what could best be described a list of declarations of intent with little practical impact (Annex 7 of the AIMDD and Annex X of the MDD), to a robust set of rules included within the provisions of the MDR, complemented by a comprehensive annex comprised of general requirements, demands on the application for clinical investigation, and a list of sponsor obligations (Articles 61-82 and Annex XV of the MDR). Concerns with setting higher standards for the safety of MDs are depicted by the provision that clinical data sourced from studies or reports must be supported by their publication in peer-reviewed scientific literature. Additionally, under the MDR, data used for showing safety and/or performance should be sourced from clinically relevant information coming from post-market surveillance – in particular the post-market clinical follow-up, in itself a reflection of the intensified post-market surveillance requirements for manufacturers – and include clinical benefits of the device when used as intended by the manufacturer.

The MDR further introduces detailed and strict criteria for the designation of NBs, which are subject to further control at EU level (recital (50) of the MDR). The MDR requires NBs to be certified under its rules (Article 36 and Annex VII of the MDR), irrespective of having been designated and certified in accordance with the directives. It also determines that certificates issued by those NBs, albeit remaining valid until the end of the period indicated in the certificate, shall become void at the latest on 27 May 2024 (Articles 38 and 120 of the MDR).

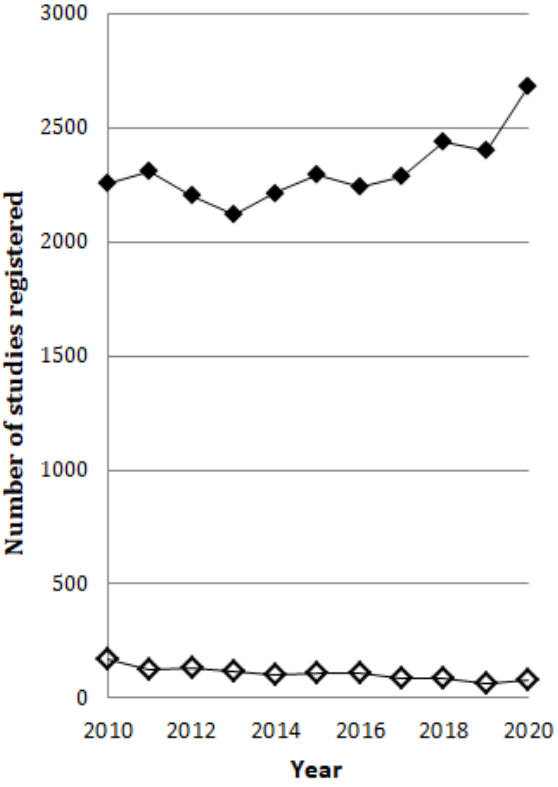
Although there are signs of an increasing matching between the regulatory frameworks for MDs and drugs, the analysis of the relevant legislation shows this has not yet been achieved. We noted that the MDR refers to clinical investigations – defined as systematic investigations involving one or more human subjects, undertaken to assess the safety or performance of a device –, whilst Directive 2001/83/EC refers to clinical trials. The improvement of electronic registries is also part of the MDR goals, as new rules determine that information on clinical investigations must be provided through an electronic system that shall be interoperable with the EU database for clinical trials on medicinal products. In this scope, information shall be accessible to the public, with the exception of information relating to clinical investigations and their revocation, suspension, termination, or modification exchanged between Member States and the Commission, and information on the application exchanged between the sponsor and the Member State concerned, unless confidentiality is justified. The MDR states that data generated from clinical investigations should be scientifically valid. Yet, in spite of the similarities

regarding the demands from generated data between clinical investigations and clinical trials – reliable and robust, taking account of statistical approach, design of the investigation/trial and methodology, including sample size, comparator, and endpoints –, the latter require randomisation of study participants, as opposed to clinical investigations (296). Additionally, we noted that the assessment of the applications for conducting clinical investigations and clinical trials contain differences that impact generated data. In fact, while an application to conduct a clinical investigation is assessed by demonstrating the investigational device(s) compliance with the applicable general safety and performance requirements, an application to conduct a clinical trial is evaluated on its relevance, especially on whether the group of subjects participating in the clinical trial represent the population to be treated. Manufacturers of MDs may, in certain circumstances, select not to resort to data from clinical investigations, relying on studies reported in the (peer-reviewed) scientific literature showing equivalence to another approved device. By contrast, pharma companies are required to provide the results of pre-clinical tests and clinical trials, unless they are requesting marketing authorisation under generic drug rules.

A summary of the regulatory standards on quality, safety, and performance/efficacy parameters for the approval of MDs and drugs in the EU and a highlight of the main differences between them is presented in Table 7.1.

Clinical research and HTA-supported recommendations for drugs and medical devices

We found a substantially larger number of phase II and III industry funded clinical studies registered in the ClinicalTrials.gov database for drugs (42,761) than for medical devices (2,039) – more than a twenty-fold difference. Additionally, when plotting the yearly studies that started over the last ten years, an upward trend was observable for drugs – with an exceptionally large increase from 2019 to 2020 (+11.7%) –, whereas a slight downward trend is noticeable for clinical trials started for medical devices (see Figure 7.3).



Number of studies started		
Year	Medical devices	Drugs
<2010	856	17,341
2010	170	2,256
2011	124	2,305
2012	135	2,200
2013	118	2,118
2014	102	2,213
2015	109	2,294
2016	107	2,240
2017	83	2,285
2018	89	2,433
2019	66	2,398
2020	80	2,678
Total	**2,039**	**42,761**

Figure 7.3 – Number of manufacturer sponsored clinical studies started for medical devices and drugs

A lower number of HTA-based recommendations was found for MDs when compared to drugs. In France, HAS issued 2,805 recommendations for drugs and 1,610 opinions for medical devices. It is further worth noting that around 10% of these opinions stated that data were insufficient for issuing a recommendation about that particular health technology. In Germany, G-BA made public 622 dossiers assessing the benefit of drugs (§ 35 SGB V assessments) and only 40 assessing the benefit of medical devices (§ 137h SGB V assessments); a smaller number was found for IQWiG (424 and 8, respectively). It ought to be said that whilst both HTA institutions are responsible for issuing assessments on the benefit of health technologies, the recommendations from G-BA are the binding ones in the scope of the formal P&R process in Germany. In Italy, there were 35 HTA reports for MDs available in the AGENAS website and 119 innovation assessment reports for drugs in the AIFA website. In Spain, AEMPS published 283 therapeutic positioning reports for drugs, while from the 630 HTA reports available in RedETS website 81 included medical devices in the scope of their analysis, from which 12 consisted of an evaluation of a single medical device. NICE – the institute responsible for issuing evidence-based recommendations developed by

independent committees in the UK – published 503 technology appraisals for drugs and 47 medical technologies guidance documents for medical devices.

The summary of the HTA-based recommendations by jurisdiction for MDs and drugs is presented in Table 7.2.

Table 7.1 – Regulatory standards on quality, safety, and performance/efficacy for the approval of medical devices and drugs in the EU

Quality		Safety		Performance/Efficacy		Main differences
Medical devices	**Drugs**	**Medical devices**	**Drugs**	**Medical devices**	**Drugs**	
• Requirements on design and manufacturing depending on the class and intended purpose of the device • Quality System: implementation of a production quality assurance system verified by a NB	• Standards on qualitative and quantitative composition • Compliance with GMP	• Risk control measures and safety principles taking into account the generally acknowledged state of the art • Compliance with design and manufacturing information • Clinical data sourced from clinical investigations or studies reported in scientific literature, published and/or unpublished reports, clinically relevant information coming from post-market surveillance • Clinical evaluation consultation procedure for by an independent expert panel	• Data sourced from clinical trials • Non-clinical (pharmaco-toxicological) studies • Pharmaceutical (physicochemical, biological or microbiological) tests	• Performance requirements as intended by the manufacturer • Clinical data sourced from clinical investigations or studies reported in scientific literature, published and/or unpublished reports, clinically relevant information coming from post-market surveillance	• Data sourced from clinical trials	• Source of data on safety and performance/efficacy for approval between MDs and drugs (clinical investigations vs clinical trials) • The MDR imposes more stringent premarket clinical data requirements, introducing that data used for showing safety and/or performance are sourced from published peer-reviewed scientific literature and from clinically relevant information coming from post-market surveillance, and demanding a summary of safety and clinical performance for Class III MDs • Clinical evaluation of MDs under the MDR imposes the verification of clinical benefits of the device when used as intended by the manufacturer

Abbreviations: EU, European Union; GMP, Good Manufacturing Practices; MD, Medical Device; MDR, Regulation (EU) 2017/745 on medical devices; NB, Notified Body.

Table 7.2 – Summary of HTA-supported recommendations in EU5 countries for medical devices and drugs

Jurisdiction	HTA institution(s)	Medical devices	Drugs
France	HAS	1,610 opinions on medical devices and other health products, from which 163 stated that data were insufficient for issuing a recommendation	2805 recommendations
Germany	• G-BA • IQWiG	• 40 § 137h SGB V assessments from G-BA • 8 § 137h SGB V assessments from IQWiG	• 622 § 35 SGB V assessments from G-BA • 424 § 35 SGB V assessments from IQWiG
Italy	• AIFA • AGENAS	35 HTA reports (AGENAS)	119 innovation assessment reports (AIFA)
Spain	• AEMPS • RedETS	• 81 (out of 630) HTA reports from RedETS include medical devices in the scope of the analysis • 12 reports are specific for a single device	283 therapeutic positioning reports from AEMPS
UK	NICE	47 medical technologies guidance publications	503 technology appraisal guidance publications

Abbreviations: AEMPS, Agencia Española de Medicamentos y Productos Sanitarios (Spanish Agency of Medicines and Medical Devices); AGENAS, Agenzia Nazionale per i Servizi Sanitari Regionali (National Agency for Regional Health Services); AIFA, Agenzia Italiana del Farmaco (Italian Medicines Agency); G-BA, Gemeinsamer Bundesausschuss (Federal Joint Committee); HAS, Haute Autorité de santé; HTA, Health Technology Assessment; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care); NICE, The National Institute for Health and Care Excellence; P&R, Pricing and Reimbursement; RedETS, Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (Spanish Network of Agencies for Assessing National Health System Technologies and Performance) SGB, Sozialgesetzbuch (Social Code Book); UK, United Kingdom.

Discussion

This study aimed at reviewing and comparing the legal framework in the EU for the approval of MDs and drugs. Additionally, it sought to compare the available information on the clinical research, as well as the HTA-based recommendations for both MDs and drugs in the EU5 countries. It was deemed that a specific analysis of the approval of MDs and/or drugs in the EU5 countries was not required, as the proper functioning of the EU single market and the free movement of goods is better achieved through centralised EU legislation that harmonises standards for MD and drug approval across Member States and that allows for relevant stakeholders to efficiently function within the EU, fostering a harmonised regulatory environment that ultimately benefits EU patients (278, 297).

Following the results on the regulatory framework of medical devices and drugs in their quality, safety, and performance/efficacy dimensions, it is critical to point out the different requirements between the clinical data necessary for the approval of MDs vs. drugs. In fact, the lower standards for MD approval are a plausible explanation for the significantly lower number of manufacturer sponsored clinical studies observed for MDs when compared to drugs (see Figure 7.3). As a consequence, comparative evidence generation seems to be scarcer for MDs than for drugs, including for high-risk devices, which are often approved without rigorous clinical studies (298). Although figures for device approvals in the EU are not available due to insufficient transparency (299), in the United States (US), even when going through the most stringent regulatory pathway for high-risk devices, MDs are typically approved based on a single non-randomised clinical study without a control group (300, 301). Between 2000 and 2011, less than half of studies supporting Food and Drug Administration (FDA) approval of high-risk cardiovascular devices included active comparators (302). As previously noted, it is expected that the MDR will provide some clarity regarding the clinical evidence used in the approval of MDs in the EU through the public disclosure of data generated in clinical investigations imposed by the new regulation. However, besides the issue of evidence generation for approving the use of MDs, there is still the need to consider the scarcity of comparative evidence, which may, in turn, impair the correct assessment of MDs within any given healthcare system.

HTA refers to the systematic assessment of the properties, effects, and/or impacts of health technology with the main purpose of informing a policy decision making (3). Unlike the regulatory approval of MDs and drugs in the EU, technology assessments are done at a country level, as they are mostly used for P&R purposes. Despite the recent political agreement on the HTA Regulation reached by the European Parliament and the Council that will enable joint scientific assessments of treatments

and MDs at EU level (303), it is not anticipated that the Regulation will impact on Member States' responsibility for the management of their health services. Hence, P&R decisions are expected to continue to be a matter of national competence of the Member States. A previous study analysing HTA and reimbursement processes in countries taking part in EUnetHTA found that, from 58 agencies in 29 countries, 94% of countries reported assessing pharmaceuticals and 68% of countries reported assessing non-pharmaceutical health technologies (304). In spite of the percentage reported by the country HTA agencies, it is critical to reflect on the relative numbers of assessments performed and published for MDs and drugs. On that topic, we found that a substantially lower number of published HTA reports or recommendations for MDs than for drugs in all analysed jurisdictions. Noteworthy, HAS in France reported that data were insufficient for issuing a recommendation for 10% of the assessed MDs, whereas this was never a problem identified for drugs (see Table 7.2). It is also meaningful to reflect upon the issue of economic evaluation, one of the main components of technology assessments. Although the general methods for conducting economic evaluations are well established, most guidelines were written with drugs in mind and they typically rely on randomized controlled trial (RCT) data for the assessment of relative treatment effect (305). MDs can therefore face hurdles in their fair assessment, seeing that their clinical trials are often non-randomized, non-blinded, do not have active control groups and lack hard endpoints (270). Drummond et al. (306) identified six reasons for why devices are different from drugs when it comes to economic evaluations: (i) many devices are diagnostics, (ii) challenges undertaking RCTs, (iii) efficacy often depends not only on the device itself, but how it is used, (iv) wider economic implications vs. drugs (e.g. training or organizational context impacting the effectiveness of the device), (v) equivalent clinical evidence may not be available for all products, hindering comparisons, and (vi) prices are much more likely to change over time because of the market entry of new products. Despite these challenges and the differences in the mandates and competencies of individual HTA agencies, additional data generation may be requested for coverage and reimbursement, as the regulatory framework should not be accepted as an argument for using lower levels of clinical evidence in technology assessments and decision making (275). Specific guidelines and recommendations for conducting technology assessments for MDs have since been published (307-309).

Some limitations of this study should be pointed out. Although we aimed at assessing the implications that the regulatory framework may have on clinical and HTA research, we did not have any relevant quantitative data for establishing a causality effect between one and the other. Moreover, we simply quantified the comparative research output for MDs and drugs, while no judgment was made about the quality of that research. Also, HTA methodology typically uses all best available evidence and not only industry funded phase II and phase III trials. For MDs in particular, HTA often

uses evidence from non-manufacturer studies that are not typically phase II and III. Although there are some limitations through searching only ClinicalTrials.gov database for medical device evidence, this approach also shows the lack of investment in developing primary evidence for the approval of MDs.

From the performed analysis, it was possible to identify some points for improvement and, as such, to issue some policy recommendations that could promote an integrated evidence-based assessment system that could result in a more efficient allocation of resources in the healthcare systems.

First and foremost, the classification of MDs. In the Results section we elaborated on the classification of MDs in regulatory terms and we aimed at defining and confining our analysis to what we labelled as interventional medical devices. In our view, these are the MDs that best match drugs and their purpose. Those MDs could benefit from a similar environment to that of drugs. While the classification of MDs from an HTA perspective will not be an easy nor consensual task, work has already been made on that topic (310).

The regulatory processes for MDs should be more closely aligned to the HTA framework. For instance, a MD that would be used for the treatment of a disease, injury or disability in the same target population as a drug (e.g. heart failure or diabetes) should be evaluated on the same outcomes. For that to happen, clinical studies aimed at generating evidence for devices should be forced to do so implicitly. The best way for achieving this goal is through a mandatory technology assessment of interventional devices (or a subset), similarly to what happens for drugs in many countries. The manufacturer would then be forced to think about the intended purpose for the MD from a clinical perspective and how that same device would prove its worth. Is there any clinical development of an oncology drug that would even risk not collecting evidence on the overall and progression-free survival? The assessment of the MD could not only serve the purpose of deciding upon reimbursement, as it could also impact the price of the MD. After all, this is what happens with drugs and what is considered everywhere to be the most efficient way of allocating healthcare resources.

Acknowledging the difficulties of the HTA of some MDs, conditional coverage and post-approval evidence development should be put in place. In a way, the MDR already foresees it, although very much focused on the safety dimension. Conditional coverage would imply the establishment of high quality registries that would entail periodic revaluations over time. Real world evidence should become a fundamental source of data for the assessment of MDs. The creation of high quality registries would serve more than the purpose of revaluating MDs, as the data contained in those

registries is frequently used in other areas of research aimed at informing decision making.

In summary, we are calling for a conceptual change in the way MDs are perceived in the healthcare sector, especially when they could fall under the umbrella of interventional devices. There is an urgent need for seeing interventional MDs as treatments and, as any other treatment in the healthcare space, assess them with the best available methodology. An integrated path for the life cycle of MDs, similarly to what happens with drugs, needs to be designed and validated in the upcoming years.

Appendices

Online Appendix 7.1

Available at:

https://drive.google.com/file/d/1cRo6lpbtDamQ9hEtlEuLgmXqo4bVfeUW/view?usp=sharing

Chapter 8

General discussion

Chronic diseases are responsible for a high disease and economic burden and they are the main cause of mortality worldwide (2, 4-7). In keeping with the World Health Organisation's recommendation to delineate comprehensive strategies for reducing death from chronic diseases more effectively, many healthcare systems started looking into digital technologies as a solution for managing chronic diseases and thus improving public health (8, 57). Early warning systems – timely surveillance systems that collect information on diseases in order to anticipate health deterioration and to trigger prompt clinical intervention – are considerably enhanced by the technological developments in recent years and they are seen as a promising solution for effective chronic disease management (55, 56). Modern healthcare systems often resort to health technology assessment methods to systematically evaluate the properties, effects, and impacts of new health technologies for informing policy and decision making (3, 72, 73). Economic evaluation – one of the critical components of health technology assessment – aims at providing evidence that rationally helps to decide upon healthcare resource allocation within a *value for money* framework (81).

The aim of this thesis was to investigate the methodology used in the economic evaluation of early warning systems for chronic disease management and to contribute to the field by developing a more generic and versatile health economic model for assessing the cost-effectiveness of early warning systems.

Thesis pathway and rationale

What exactly are early warning systems?

An overarching definition of early warning systems (EWS) was required in the scope of this thesis. The United Nations define EWS in the context of Climate Action as *an adaptive measure for climate change, using integrated communication systems to help communities prepare for hazardous climate-related events* (311). A quick search on a well-known Internet browser shows that the top results returned for EWS are related to geohazards such as landslides, volcanoes, and earthquakes. Other scientific areas also use the term EWS: a paper in the field of international monetary economics/international finance – ironically published in 2006 – presents the development of a new EWS model for predicting financial crises (312). Digging deeper into the EWS definition unveiled a publication from the United Nations Development Programme, which identifies four key components of EWS: risk knowledge, monitoring and warning services, dissemination and communication, and response capability (313). These elements are the pillars of the EWS concept, regardless of the point of view from which it will be studied.

When translating these elements to the field of health care, after acknowledging a pre-identified risk, we conceptualised EWS in this thesis by adapting the three remaining

three main elements and defining them as such: (i) monitoring and collection of clinical data (e.g. vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a particular combination of signs and symptoms – that trigger an alarm and follow-up actions (59). Although this is not intended to be a clear-cut definition of EWS, any health technology comprising the three previously identified elements was interpreted as an EWS in the context and throughout the thesis.

Published decision analytical models for the economic evaluation of early warning systems

The first step of the research consisted of investigating the state-of-the-art of decision analytical models (henceforth simply referred to as models) for the economic evaluation of EWS for chronic disease management. For that purpose, a systematic literature review of the references catalogued in nine key electronic databases was performed. None of the searched databases included the exact expression "early warning systems" in their thesaurus terms. Therefore, considering the interpretation given to EWS in the context of the thesis, in order to ensure that no relevant studies were being missed, a sensitive search strategy describing EWS through a wide variety of related search terms was adopted (see Appendix 2.1). This search returned a large number of references with a great deal of false positives, i.e. studies that did not match the inclusion criteria but were identified by the broad search strategy. For that reason, the scope of the review had to be narrowed from *chronic diseases* to *chronic heart failure*, lest the amount of retrieved references would become unmanageable. Heart failure (HF) was the chosen disease, as a result of the available access to the database from the Trans-European Network-Home-Care Management System (TEN-HMS) study (314) and the outlook of using those data in subsequent analyses included in the thesis.

Chapter 2 presents the systematic literature review of cost-effectiveness models for the management of chronic heart failure. The study described the general and methodological characteristics of the reviewed models. The results were reported in accordance to the framework defined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines published in 2009 (109) – as these were in force at the time of publication of the review included in the thesis –, although a recent update of the PRISMA guidelines has been published in 2021 (315).

A quality assessment of the methodological characteristics of the included models was performed using the checklist for the critical appraisal of decision analytic models for

health technology assessment developed by Philips et al. (110). This assessment helped to determine that models were unsatisfactory in the consideration and discussion of any competing theories regarding model structure and disease progression, the identification of key parameters and the use of expert opinion, and the assessment of the four types of uncertainty (methodological, structural, heterogeneity, and parameter). Describing the different approaches used in published models also showed that the most frequently used methodological approaches in decision analytic modelling – including the model type: decision trees and Markov models – did not seem adequate for the assessment of EWS. These conclusions provided useful insights for the development of the model presented in **Chapter 5** of the thesis.

Critical events: prediction and impact

By definition, the relevant health technologies in the scope of the thesis provide an *early warning* for an occurrence, which is normally a critical event that patients are at risk of experiencing and that should be predicted and avoided from a clinical perspective. It is vital, therefore, to identify the event(s) of interest in each of the clinical situations being investigated.

In the instance of HF, one can think of a myocardial infarction as the event of interest, while in chronic obstructive pulmonary disease patients we may be interested in predicting and avoiding exacerbations, which normally result in negative health outcomes and increased costs (316, 317). However, when considering that the focus of the thesis is the management of chronic diseases, from a broader perspective, we can also think of an event of interest such as hospitalisation, transversal to most chronic diseases, and that is desirably avoided in the management of chronic disease patients who often have numerous comorbidities and that tend to be of older age.

Following on the considerations above, **Chapter 4** describes the concept of a diagnostic algorithm (DA) for estimating the risk of HF-related hospital admissions using an ambulatory telehealth programme that combined clinical software and in-home remote monitoring technology (the EWS under analysis). The algorithm intends to provide healthcare professionals with a global risk score and, ultimately, using that score to define follow-up actions based on evidence-based thresholds of the risk of hospitalisation.

Hospitalisation in HF is a consequence of acute decompensated heart failure, a clinical condition characterised by a rapid onset or worsening of symptoms or signs of HF that requires urgent evaluation and treatment and that typically leads to an urgent hospital admission (10-12). Since the risk for hospitalisation depends on the individual disease and patient characteristics (34, 35), it is precisely where the

previously introduced concept of an early warning system comes into play. Looking retrospectively at the data generated by the EWS under analysis, which contained daily measurements of patient and disease characteristics (body weight, systolic blood pressure, diastolic blood pressure, heart rate, and blood oxygen saturation and answer to surveys administered remotely on oedema, fatigue, shortness of breath, and activity status) and the information on hospitalisation events, it was possible to develop an algorithm that could predict hospitalisations based on the patient and disease characteristics. The algorithm is based on a series of different logistic regression models that use the different characteristics (e.g., vitals, surveys, etc.) for calculating the risk of being hospitalised. The logistic regression models are combined in a final risk score by taking their average of their results. This was referred to as an ensemble algorithm, which allowed for overcoming the main constraint found in using a simple logistic regression model with the available dataset, i.e. the need to feed the model with a complete feature vector in order to generate a risk score.

Another critical feature of early warning systems is the existence of an alarm when the calculated probability of hospitalisation goes above a certain risk score threshold. In this framework, we can interpret the alarm as a diagnostic test: if an alarm is raised (above the threshold), the test is positive; if not (below the threshold), the test is negative. We can then consider the event of interest (hospitalisation) as "having disease/condition" and not being hospitalised as "not having disease/condition". Table 8.1 presents the confusion matrix for the described situation.

Table 8.1 – Confusion matrix for alarm and hospitalisation

	Hospitalised	**Not hospitalised**
Alarm	True positive	False positive
No alarm	False negative	True negative

From the confusion matrix above, type I (false positive) and type II errors (false negative) are clearly defined, which allows us to calculate the statistical measures of the performance of the algorithm: sensitivity (also known as true positive rate, recall, or probability of detection) and specificity (also referred to as selectivity or true negative rate). As any other diagnostic test, sensitivity and specificity will depend on the risk score threshold used for the algorithm and they are inversely proportional between themselves (i.e. as sensitivity increases, specificity decreases and vice versa). Plotting sensitivity against *1 – specificity* at various thresholds results in a receiver operating characteristic (ROC) curve.

Depending on the probability of being hospitalised calculated by the algorithm and the consequential alarm, it is possible to conceive follow-up actions/pathways. Although

it was not the goal of the development of the algorithm to assess the optimal follow-up actions after an alarm, an exemplificative diagram of possible pathways after an alarm is raised by the EWS is shown in Figure 8.1.

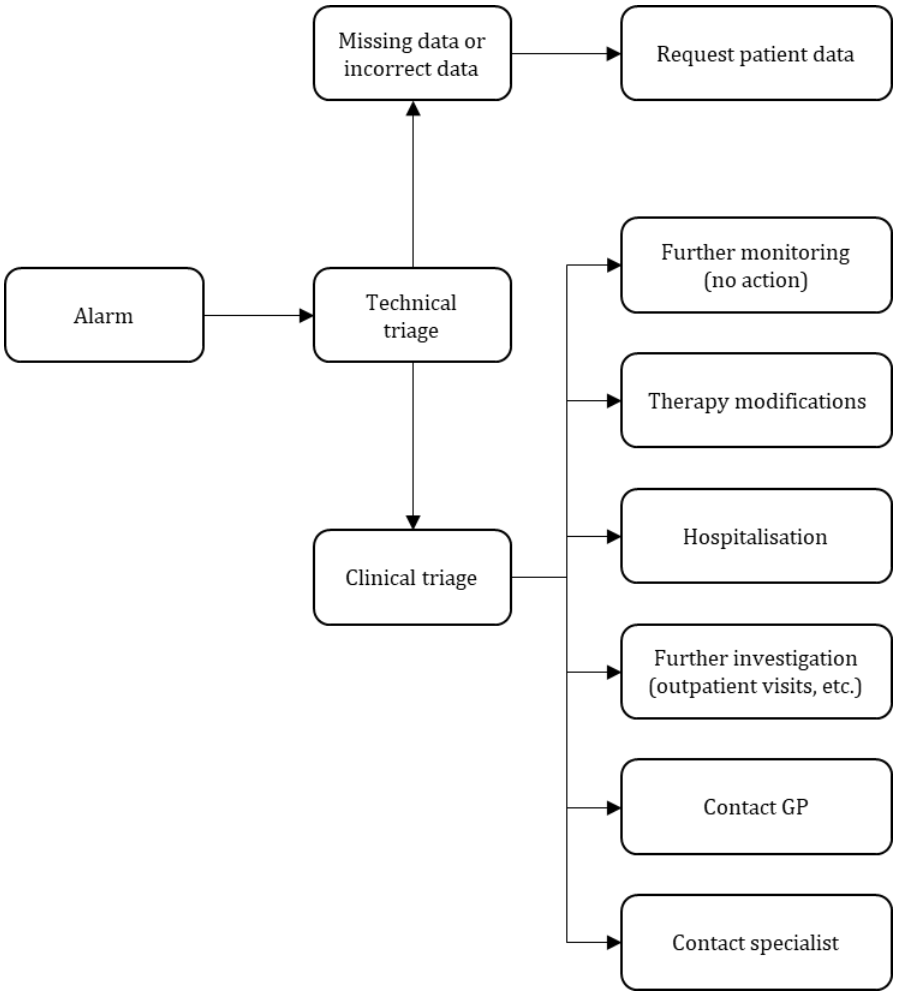


Figure 8.1 – Follow-up pathways of an alarm in an early warning system

In regards to the determination of the correct operating point for the alarm, **Chapter 4** uses the therapeutic threshold concept (177) for calculating the risk score threshold for the probability of being hospitalised above which an alarm should be raised. When applying the monetary value of a planned hospitalisation and an unplanned hospitalisation as the measurement of the value of the outcomes (using the length of hospital stay as a proxy for distinguishing between hospitalisation types), a patient should be followed-up after an alarm when his/her probability of being hospitalised,

as calculated by the algorithm, is higher than 11.7%. However, it should be stresses that using another measurement for the value of the outcome (e.g., quality of life) could lead to a different result for the correct operating point for the alarm.

For a full assessment of the added value of the EWS, we ought to evaluate the entire package of interventions, including the home telemonitoring system, the diagnostic algorithm, the cut-off value for the alarm to go off, and the clinical intervention that follows the alarm. Consequently, a full cost-effectiveness analysis that can properly assess the costs of implementing and using the technology and the anticipated effectiveness resulting from the early detection of disease deterioration and the subsequent reduction of hospitalisations is required. The conceptual modelling approach to such a full cost-effectiveness analysis was described in **Chapter 5**.

With a full cost-effectiveness analysis in mind, considering hospitalisation avoidance as one of the main effects of early warning systems, the impact of hospitalisations on health-related quality of life (HRQoL) in patients with chronic heart failure was explored in **Chapter 3**. The impact of nonfatal hospitalisations on the HRQoL for a cohort of patients previously diagnosed with heart failure was estimated by calculating the difference in utility measured using the EQ-5D-3L in patients that were hospitalised and had records of utility before and after hospitalisation. The mean difference between HRQoL measurements pre and post hospitalisation was found to be 0.020 [95% CI: -0.020, 0.059], when measured with the EQ-5D index, while there was a mean decrease of -0.012 [95% CI: -0.043, 0.020] in the utility measured with the visual analogue scale. Differences in utility according to the primary cause for hospitalisation were identified: hospitalisations due to respiratory/chest infection and ventricular tachycardia showed an improvement in quality of life when considering the index utilities measured before and after admission, while hospital admissions attributed to atrial fibrillation and myocardial infarction showed a decrease between index utilities measured before and after hospital admission.

The model: a discrete event simulation approach

The systematic literature review presented in **Chapter 2** revealed that published models had considerable drawbacks for the assessment of the cost-effectiveness of early warning systems for heart failure management, which can be extrapolated to the broader concept of chronic disease management. More specifically, decision trees and Markov models – the model types used in all identified studies – failed to capture a critical aspect in chronic disease management done with early warning systems: the impact of individual patient and disease characteristics on the outcomes.

Chapter 5 presents the construction and validation of a discrete event simulation (DES) model that is able to model heart failure patients managed with usual care or an

EWS (with or without a diagnostic algorithm) and to account for the impact of individual patient and disease characteristics on their outcomes. The model was developed using patient-level data from the TEN-HMS study, coded using the programming language R, and validated along the lines of the Assessment of the Validation Status of Health-Economic decision models tool (AdViSHE). The model includes 20 patient and disease characteristics and generates 8 different outcomes.

The use of DES proved to be a suitable approach for modelling the analysed patient population. That type of models, in allowing for individual patient events to be recorded, is generally appreciated by decision-makers in their appraisal of economic evaluations. First, because creating a patient history simplifies the explanation of the general lines of modelling to people who are not specialists in decision analytical modelling itself but are familiar with the output generated by these studies. And second, because DES allows for the exploration a variety of hypothetical scenarios, which is sometimes impossible using the more common Markov models and decision trees. A frequent problem in economic evaluations is the rigidity of the models and the inability of the developer to change them readily to answer "what if" questions. On a different note, it should be mentioned that the flexibility of the programming software amply facilitates the adaptation and debugging of the model.

However, DES models also have their perils. The most flagrant of them all is arguably their data requirements. Primarily, the need of patient-level data: even when those data exist, they are frequently incomplete or of poor quality, which generally leads to the need for some assumptions regarding the data or to the exclusion of patients with missing values, leaving the analyst to face some dilemmas on how to use the data at disposal in the best possible way. A more comprehensive examination of the strengths and limitations of DES models was outlined in the Discussion section of **Chapter 5**.

The health economic model was developed using RStudio, an integrated development environment for the programming language R. Both R and RStudio are free and open-source software for data science, scientific research, and technical communication, with the mission *to enhance the production and consumption of knowledge by everyone, regardless of economic means, and to facilitate collaboration and reproducible research, both of which are critical to the integrity and efficacy of work in science, education, government, and industry*. As previously noted, this thesis had the goal of developing a generic and versatile health economic model for assessing the cost-effectiveness of early warning systems. Using R software allowed us to build a model that embodies that aspiration, both in terms of the adaptability and the availability of the model for interested parties: adaptability to other diseases and early warning systems, by changing the critical events for the specific disease without changing the core model structure; availability through making the code available to others in a development

environment that is open source and free-of-charge (https://github.com/fernandoalbuquerquealmeida/EWS_HF_DES_model).

Cost-effectiveness results using the model

Chapter 6 uses the model described in **Chapter 5** for assessing the cost-effectiveness of a home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands. Three interventions were included in the cost-effectiveness analysis: (i) usual care (UC) – patient management plan implemented by the patient's primary care physician (314), (ii) HTM (as described in the TEN-HMS original publication (314)), and (iii) HTM with the addition of a DA (HTM+DA).

The Dutch guidelines for economic evaluations in healthcare require a societal perspective, including all costs inside the healthcare sector, patient and family, and other sectors, regardless of who is paying for those costs, productivity losses assessed using the friction cost method, and future unrelated medical costs (210).

The base-case analysis revealed that HTM+DA extendedly dominates HTM and it has a deterministic incremental cost-effectiveness ratio (ICER) versus UC of €27,712 per quality-adjusted life year (QALY). The scenario where a healthcare perspective was taken resulted in an ICER between HTM+DA and UC of €14,408/QALY, which shows that many of the costs resulting from increased life expectancy fall outside of the healthcare sector although they take an economic toll on the wider society.

The cost-effectiveness threshold in the Netherlands depends on the burden of disease as measured by the fraction of QALYs that people lose relative to the situation in which the disease had been absent – referred to as proportional shortfall. The appropriate cost-effectiveness threshold for the analysed patient population was €80,000/QALY. At this threshold, probabilistic sensitivity analysis determined that HTM+DA had a 96.0% probability of being cost-effective.

The guidelines for economic evaluations in the Netherlands require the calculation of the expected value of perfect information (EVPI) when the probability that the intervention is cost-effective at the appropriate cost-effectiveness threshold is lower than 100%. The calculated EVPI per patient was €341. With an estimated number of eligible patients for the HTM-based interventions in the Netherlands of 253,118 patients for the period 2020-2024, the population EVPI was estimated at €86,383,575 (after discounting). Two main things should be pointed out in this regard. First, we did not have the standard error for many of the variables included in the model and we assumed 10-20% of the mean for that value, which, despite being usual in cost-effectiveness analyses, is a procedure that artificially imposes the extent of the impact of parameter uncertainty in the model results. Second, the deterministic incremental net monetary benefit (NMB) of HTM+DA versus UC at a €80,000/QALY was estimated

at €34,207. Thus, the EVPI corresponds to circa 1% of the NMB of implementing the HTM+DA intervention, which can be seen as quite a low risk of making the wrong decision due to uncertainty.

The ability to include the effect of a diagnostic algorithm in the overall assessment of the cost-effectiveness of the EWS through the use of the figures for sensitivity and specificity is one of the innovative characteristics of the model presented in **Chapter 5**. But while the values for sensitivity and specificity give an indication of the accuracy of the diagnostic algorithm as a diagnostic test, they do not provide any information about the operating point at which they result in the best cost-effectiveness ratio for the intervention as a whole. The model allowed for calculating the cost-effectiveness of the HTM+DA intervention at different operating points of the ROC curve through using different combination of sensitivity and specificity or, in other words, by weighting the relative costs and effects of the false positives and the false negatives. But since sensitivity and specificity are not independent variables, the optimal operating point within the ROC curve ought to be found iteratively, as there are a great deal of calculations running in the background of the model and it would be mathematically very hard to solve all those equations for sensitivity and specificity. Further, in spite of using a patient-level model, the cost-effectiveness of the algorithm in the model is included as the average of the population, as it uses the figures of sensitivity and specificity of the whole population and does not calculate the hospitalisation risk for each individual simulated patient. Increasing the sensitivity of the algorithm by setting a lower threshold for the alarm to go off, which entails an increase in the false positive rate (decreased specificity), resulted in a higher number of avoided hospitalisations, life years, and QALYs, but with higher total costs. Contrarily, decreasing the sensitivity (i.e. setting a higher threshold for the alarm) resulted in lower costs and worse health outcomes. From the previously described exercise, the model proved suitable for analysing the cost-effectiveness of the HTM+DA intervention for each combination of sensitivity and specificity within the ROC curve of the diagnostic algorithm that was used in the analysis.

Another key feature of the DES model is the ability to assess the cost-effectiveness of specific subgroups of patients with relative ease. The model can be used for performing subgroup analyses by defining subgroups of interest in the whole patient-level database by creating a model population through randomly sampling patients from that subpopulation. We analysed a wide range of subgroups and we found that the subgroup of NYHA class IV patients was the one that recorded the highest deviation from the base-case analysis results, with an ICER of €52,727/QALY (+90.3%) for HTM+DA vs. UC. Conversely, the subgroups with the better cost-effectiveness ratios were patients younger than 65 years of age and patients belonging to NYHA class I (€22,830/QALY [-17.6%] and €22,870/QALY [-17.5%], respectively).

In respect to the subgroup analyses, as discussed in greater detail in **Chapter 6**, it is essential to highlight that the interpretation of subgroups in patient-level simulation models is not entirely straightforward, as every subgroup created from the baseline population by restricting the values for the variables of interest is likely to generate differences in the remaining patient and disease characteristics when compared to the model population used in the base-case analysis. Thus, the observed changes in the cost-effectiveness versus the base-case results are not only due to the characteristic controlled in the subgroup (e.g., age), but also from the other patient and characteristics that may be associated with it (i.e. older patients may also have other comorbidities like diabetes or chronic obstructive pulmonary disease or belonging to more advanced stages of disease burden like NYHA class IV).

The available body of evidence for early warning systems

The economic evaluation of health technologies implies having access to a great deal of data. Not only the data generated from clinical trials, but also the data from observational studies that may need to be generated for determining the effectiveness or safety in a real world setting. Health economic models are a complicated mathematical framework for performing cost-effectiveness analyses that aim at including all relevant inputs and make use of all the available body of evidence for estimating the value and uncertainty around those inputs.

Early warning systems are normally medical devices, especially when seen from a technological point of view. Medical devices are traditionally less regulated than drugs and the standards of evidence generation and collection for their market access are generally lower. Having access to lesser quality data or not having access to those data at all constitutes impairment for the accurate assessment of the cost-effectiveness of medical devices (i.e. EWS).

Chapter 7 presents a review and comparison of the legal framework for the approval of medical devices and drugs in the EU and it presents the available information on manufacturer sponsored clinical studies and HTA-supported recommendations for medical devices and drugs in those same jurisdictions.

Since the approval of medical devices and drugs in the EU is centralised (exceptions do exist for drugs, but they are not relevant in the context of this analysis), the standards for the approval of devices and drugs in terms of their quality, safety, and performance (term used for medical devices) / efficacy (term preferred for drugs) were analysed using EU legislation.

HTA-supported recommendations for devices and drugs are normally issued by national agencies in the EU, as these are normally used for pricing and reimbursement purposes, which are decided at a member state level and not centrally. Thus, the HTA-

supported recommendations from the HTA agencies of the EU5 countries were analysed (although the UK has formally withdrawn from the EU, the UK was considered a EU5 country, as most of the data included in the study pertain to a period in which the UK was a member state of the EU).

The review of the legislation identified different standards in the requirements for the approval of devices and drugs on their quality, safety, and performance/efficacy dimensions as well as substantially lower number of manufacturer-sponsored clinical studies and HTA-supported recommendations for medical devices versus drugs and an indication of lower standards of evidence used in recommendations for medical devices. These findings substantiate some of the challenges encountered during the research performed in the scope of the thesis. Although there is a lot of information on digital technologies (electronic processes and communications, the internet, and other information technologies) – usually labelled under eHealth, or mHealth if involving mobile devices –, the publications on those topics are normally policy related and they use mainly qualitative research. The lack of clinical studies for the approval of medical devices weakens the body of evidence available for the correct assessment of medical devices in an HTA/economic evaluation framework, especially when they should be compared to drugs that have the same therapeutic goal and that are targeted at the same disease and/or patient population.

Challenges and limitations

The methods and findings of the thesis were discussed in detail in each of the corresponding chapters. This section intends to discuss the main challenges and limitations regarding the thesis as a whole and in the context of its objectives.

Scope of the model

As previously stated, the premise of this thesis was the possibility of developing a generic model for assessing the cost-effectiveness of early warning systems. This prospect was grounded on the identification of common mechanism of action of EWS – periodically measuring individual patient characteristics in order to anticipate health deterioration and to trigger prompt clinical intervention –, regardless of their target disease. However, the work that led to the development of the model and the model itself are focused on heart failure as the chronic disease and home telemonitoring as the EWS intervention. There are two main reasons for having taken this approach.

Firstly, as previously discussed, we restricted the systematic literature review to chronic heart failure, although it is likely that at the time of the review there were

published models for other chronic diseases (e.g., COPD, diabetes) that could provide valuable information for the development of our own model.

Secondly and more importantly, the availability of high-quality data is often a bottleneck in the development of valid health economic models. Since we had access to a comprehensive patient-level database of a home telemonitoring system used in the management of heart failure, it seemed obvious to concentrate our efforts in that disease and start off with a strong base for the inputs for a model that was likely to be demanding in terms of data requirements.

That said, the findings in the thesis do not contradict its initial premise and they definitely contribute to the overall objective of reaching a generic model for chronic disease management using EWS, especially when taking into account the type of model and the coding platform used for its development. The DES simulation and the fact that the model consists of code written in R allow for adapting the model to other EWS interventions and diseases, as the core model structure can remain fairly similar. The time-to-event concept of the model is adaptable to other diseases and EWS interventions, as the event of interest that is being predicted and avoided by the EWS can be changed, as well as the regression equations predicting those events. Inevitably the code needs to be changed, but the gist of the code is a good starting point for future modelling endeavours for EWS used in the management of other chronic diseases.

Data used in the model

Although we had access to the comprehensive database generated in the TEN-HMS study (314), those data had some limitations, mainly derived from the small sample size.

Given the DES nature of the model, patient-level data were a requisite. The database consisted of a total of 426 subjects, split into 168 for home telemonitoring, 173 for nurse telephone support, and 85 for usual care. Since nurse telephone support was not an intervention of interest in the context of the thesis, those patients were not used for estimating time-to-event regression equations, although the full 426 subjects were included for drawing the simulated model population – justified by the randomised nature of the trial. While the outcomes from the nurse telephone support arm were not usable in the modelling scope of the thesis, the baseline characteristics of those patients are not proven to be different from the other two arms and using them in the simulation resulted in a larger population, thus decreasing the uncertainty around the patient baseline characteristics.

Regression equations

For the estimation of the regression equations predicting time-to-death and time-to-hospitalisation we used the 14 patient and disease characteristics that were deemed significant independent predictors of mortality in HF patients in the study by Pocock et al. (33). The TEN-HMS dataset was fairly complete, but missing data in any of the 14 characteristics used in the regression equations would result in excluding the patient from the analysis (see Online Appendix 5.1). Due to the small number of patients available for estimating the regression equations, this was not an option in our study. As such, we replaced missing data with the average of the whole TEN-HMS population, although we know that those were not the actual values for those patients if they would have been registered in the database. There are other techniques used for replacing missing data in health economic studies (227-231). We did not test whether handling missing data in a different way would significantly change the results of the model.

The issue of missing data also pertains to the logistic regression equation predicting death in hospital. For that particular regression equation there is the additional issue of determining the correct covariates for the model. In our study we used age, gender, previous history of myocardial infarction and/or chronic atrial fibrillation, comorbidities (diabetes and/or COPD), and the number of previous hospitalisations. Other covariates can be argued to have an impact on in-hospital mortality. Again, due to the small number of patients – which further shrinks for the subgroup of patients that die in the hospital –, using covariates based simply on the statistical significance of the estimated coefficient can be misleading and we chose clinical reasoning over the mathematical one.

Concerning time-to-hospitalisation, there is a limitation that also stems from the low number of patients in the dataset. The time-to-first-hospitalisation may not be the same as the time-to-second-hospitalisation and so forth. However, the number of patients in the database that experience consecutive hospitalisations is ever decreasing. A decision had to be made between assuming that the time-to-hospitalisation is the same regardless of how many previous hospitalisations a particular patient has experienced or to increase the uncertainty of the estimate for the time-to-hospitalisation with every increasing hospitalisation number. We assumed that time-to-hospitalisation was the same regardless of number of previous hospitalisations, as the number of patients experiencing multiple hospitalisations quickly dropped, thereby leading to highly uncertain estimates beyond the time-to-first-hospitalisation.

Utilities

Utilities are a critical part of health economic models. The utilities used in the model were estimated elsewhere (205), in a study that also used the TEN-HMS dataset. Utilities reported in that study depended on the NYHA class of the patient, although we did not model the change of NYHA class in the simulation. Thus, since the critical event predicted in the model is hospitalisation, we aimed at determining the impact of hospitalisation on patient utility (see **Chapter 3**). The results of that study ended up not being used as an input in the model. There was evidence of utility changes due to hospitalisation, but the direction of those changes greatly depended on the reason for hospitalisation, which was not determined in the simulation. Further, the results of the impact of hospitalisation in utilities showed that there were some inconsistencies, as the TEN-HMS trial was not designed for that particular purpose and there were uncontrolled variables in the utility measurements pre- and post-hospitalisation. For instance, the utility measurements were periodical (every 4 months), regardless of the time of the hospitalisation in relation to utility measurements.

Resource use and costs

A distinction should be made between the data used for resource use and costs in **Chapter 5** and **Chapter 6**. The data used in the construction and validation of the model (**Chapter 5**) were overly simplified, as the focus was not to provide an accurate cost-effectiveness estimate but rather to assess whether the model was able to generate credible outcomes. On the other hand, the data used in **Chapter 6** had a higher level of detail and they were suited to the particular setting of the Netherlands. In that sense, the cost-effectiveness results of each chapter should be interpreted in line with the previous explanation: in **Chapter 5** they should be loosely compared to other studies for the purpose of validating the model, while in **Chapter 6** they should be usable for the purpose of decision making about the home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands.

Sensitivity and specificity: population vs. individual

The conceptualisation of the EWS and the diagnostic algorithm was explained in detail in the *Conceptualisation of early warning systems and the diagnostic algorithm for the management of heart failure* in **Chapter 5**. The inclusion of the effect of the diagnostic algorithm was done through the measures for sensitivity and specificity. It should be noted that the figures for sensitivity and specificity pertain to the population and not the individual. At a first glance it seems an inconsistency to use these figures when we are using a DES model, which has the big advantage of being able to assess the effect of patient-level characteristics on the clinical and economic outcomes generated by the model. So, it would seem logical to use the patient characteristics to calculate the probability of being hospitalised – using, for instance, a logistic regression – and to

model the patient pathway according to the calculated probability and the thresholds previously defined for taking any particular action for the patient. However, the ability to use the actual values of the patient characteristics *in real time* (at shorter time intervals, to be more precise) is not straightforwardly implementable in the model. In order for that to happen, the model should be able to predict the trends/evolution of each of the patient characteristics as regularly as the diagnostic test is performed by the algorithm. Although we tried to estimate the evolution of patient characteristics using the TEN-HMS dataset, this was not possible, as we are dealing with 14 variables, which are often correlated and for which we only had quarterly measurements. Using daily values for each of the characteristics, although conceptually desirable, would introduce noise in the model that would be difficult to explain and that could make it harder to identify the signal from the noise.

Scarcity of data sources

Besides the limitations of the TEN-HMS dataset discussed above, the lack of other reliable data sources should be pointed out. When looking for data that could inform the model, we were inevitably confronted with the lack of data concerning early warning systems and diagnostic algorithms. Notably, we were unable to find alternative patient-level datasets to validate the model against the outcomes generated using those input data. The alternative to split the data from the TEN-HMS dataset and to use one part for the model training and the other part for the model validation was impossible due to the small size of the sample, which would raise questions concerning the validity and certainty of generated results from either of the sub-datasets.

Also, ROC curves of diagnostic algorithms were not always available, even when those algorithms were described in scientific publications. We used the ROC curve published by Koulaouzidis et al. (244), although the diagnostic algorithm presented in the publication was not used for the same intervention as we investigated and they did not use the whole array of variables that we used for estimating the regression equations in our model.

That said, the challenges encountered in getting access to other data sources were the basis for the development of the hypothesis that the lower standards for the approval of medical devices when compared to drugs could be hindering the correct assessment of those devices (in this case, EWS). In fact, that hypothesis triggered the work that was presented in **Chapter 7**, where we present evidence that the hypothesis was true.

The discussion above concerns the scarcity of data specific to the model developed in this thesis. However, there are wider limitations of modelling that relate to their

inherent structure and the availability of data for describing that same structure in mathematical terms.

Model features: the inevitable dilemma of simplifying reality

In the words of George E. P. Box: "*All models are wrong, but some are useful*". Although a commonplace, this statement could not be closer to the truth. The reason why all models are wrong is the inevitable need for simplifying reality, i.e. to focus the model on its key factors when, in fact, we are fully aware that we are disregarding other factors that may have an impact on the outcomes we are trying to simulate. From a pragmatic approach, model development and the features included in the model present a trade-off between accuracy (loyal representation of the reality) and technical feasibility. The feasibility issues often derive from the lack of adequate data to model some phenomena in a mathematical framework.

Working on a DES framework, it is easy to envision a super model that simulates every step of the patient pathway from the moment he/she enters the model until his/her death. For instance, our model could include equations for predicting the change over time for every patient and disease characteristic, which would result in an update at every event. Based on the updated characteristics, medication could be changed, which would not only influence the change in the patient characteristics but also the probabilities for experiencing a future event. Continuously (or, at least, daily) updating patient characteristics would allow calculating the risk score for the individual patient and determine whether it would exceed the predefined threshold (which could also be individualised on account of patient history and correlation between repeated measurements) and if an alarm should be raised. Based on the risk score and individual patient characteristics, a plethora of follow-up actions could be simulated (see Figure 8.1), each with different consequences, such as changes in the medication or changes in the risk threshold for that patient. These are only the most obvious features that are not part of the model but that would make perfect sense from a clinical standpoint; one could keep going about all the possible pathways for HF patients. Further, although we focused on hospitalisation as the key event in HF modelling, in other disease areas multiple events may be relevant.

So, why not including the above features in the model? The simple answer is "data and computational constraints". Although those features would accurately represent reality, it would be virtually impossible to find reliable data to accurately inform the model on them. By overcomplicating the model we would be forced to make too many assumptions, greatly undermining the validity of the results. In addition, running all the calculations for the hypothesised features above would be extremely demanding in computational terms. For instance, the daily computation of the risk score, the logical decisions the model would have to take following the results, the updated of

patient characteristics, and so on would have to run 365 times for every year a patient would survive in the model.

From the discussion above, it is crucial to extract that the trade-off between accuracy and technical feasibility should be found in the principle of parsimony. The model should be as complex as needed in order to be informative for helping decision makers in their ungrateful task of making decisions.

Decision making using economic evaluations in health care

We have discussed some of the limitations of economic evaluations and the implications they may have in the results obtained from those studies. Making decisions in health care using results of economic evaluation should take into account the points made in the previous discussion. Although economic evaluation methodologies have evolved significantly since they were first introduced in decision making in health care, the intrinsic problem concerning their scope did not change significantly. Economic evaluations use a cost-effectiveness logic that generates results in cost per QALY gained (captured by the ICER). Under this framework, QALY gains using the new intervention and the costs associated with generating these gains are compared to a reference case, which is normally the current standard of care for the problem under analysis.

Economic evaluations of new health technologies are normally conducted from a wider societal perspective or a narrower healthcare perspective (81, 87). Decisions are made upon the monetary investment that needs to be made in order to generate an extra QALY. From a societal perspective, the underlying goal is to maximise social welfare from the healthcare budget and, as such, the allocation of resources towards a new health technology is done when its ICER is lower than the threshold value specified in terms of the societal willingness to pay for a QALY; in other words, when the new health technology is considered welfare improving (87, 318). From a healthcare perspective, the goal consists in maximising population health from a (fixed) healthcare budget. In this setting, the monetary value of a QALY represents the opportunity costs of resource allocation decisions within health care (87, 318, 319).

However, there are many more criteria that should matter for decision makers when setting healthcare priorities. Decision makers should go beyond cost-effectiveness considerations and include other concepts in the overall value of a new health technology, such as equity, affordability, innovation, disease burden, or others that may be valued by society. In fact, those who work in the field are aware that other criteria besides cost-effectiveness are used for making decision, many times in a non-systematic manner.

An emerging alternative for overcoming the restricted framework of typical economic evaluations is Multi-criteria Decision Analysis (MCDA). Keeney and Raiffa first defined MCDA as "*an extension of decision theory that covers any decision with multiple objectives. A methodology for appraising alternatives on individual, often conflicting criteria, and combining them into one overall appraisal…*" (320). So, MCDA supports decision making by allowing for a systematic trade-off between multiple, sometimes conflicting, criteria. The trade-off is done in an explicit, transparent, and consistent way through establishing preferences between alternative interventions by reference to an explicit set of objectives for which measurable criteria are established (321). In the case of early warning systems, one can think about criteria that are not included in traditional economic evaluation, such as the increased feeling of security for stemming from frequent monitoring of their health status, the additional comfort of being followed at home, or the release of specialised healthcare professionals for other tasks that may be short on personnel. All these criteria and others, which may be important for deciding upon the uptake of early warning systems, are not captured under a QALY framework.

The extent to which the criteria are fulfilled is the basis for making decision for preferred alternatives and, as consequence, setting healthcare priorities. Recent literature advocates that MCDA, divided into three main categories – qualitative MCDA, quantitative MCDA, and MCDA with decision rules –, despite having a large potential to support HTA agencies in setting healthcare priorities, should see its implementation improved (322). The authors further argue that HTA agencies should include a deliberative component in their process of formulating recommendations and that they should report these deliberations, including the considerations underlying a recommendation in order to ensure the consistency and transparency of recommendations. Further, the authors recommend agencies, at a minimum, to undertake qualitative MCDA, i.e. using explicit criteria as a way for improving the quality, consistency, and transparency of recommendations, which is indisputably better when compared with employing no specific method at all.

Decision making will always be cursed by uncertainty. One of the main issues of nowadays decision making is the seeming need to substantiate every decision with quantitative data, as if quantification would shield us from the inherent uncertainty of the world. As previously noted, using data, although fascinating in its nature and from a human development point of view, is covered in extensive challenges and limitations. We should not fall into the trap of using bad quantitative studies in detriment of logical and rational arguments based on good rhetorical principles.

Implications for different stakeholders

The findings of this thesis have underlying implications for various stakeholders and at different levels of the healthcare system.

Health economic researchers

The model presented in **Chapter 5** is a unique patient-level simulation model that can be used for simulating a wide range of outcomes for different patient subgroups and treatment scenarios in HF patients. It provides useful information for guiding research and for developing new treatment options by showing the hypothetical impact of these interventions on a large number of important heart failure outcomes. Although the model was developed especially for assessing early warning systems for the management of HF patients, its core structure allows for adapting it to other diseases. We have also developed the model in R and we have made it publicly available to anyone who may be interested in the topic (https://github.com/fernandoalbuquerquealmeida/EWS_HF_DES_model). Our approach to the problem set in the thesis was from a purely collaborative standpoint, i.e. there was a preoccupation that the work we did could be useful for others and not another black box model. In this way, any researcher with access to data concerning the problem of EWS for chronic disease management can save some time and obtain ideas from the modelling work that we have performed.

Clinicians

From the point of view of the clinicians, the findings of this thesis suggest an improvement of health outcomes when using home telemonitoring systems in the management of heart failure, especially when a diagnostic algorithm is added to the intervention.

The Dutch setting was the one used for the analysis and the transferability of the results to other jurisdictions has not been assessed in this thesis. However, the assumption that the clinical inputs for the model are valid for the Netherlands, underpinned by the fact that one third of the patients included in the TEN-HMS trial were treated in the Dutch setting, could also be applied to Germany and the UK – the other two countries where the trial took place and that also included approximately one third of the remaining patients each.

Although it should be subject to a more careful analysis, we can hypothesise that the clinical results can also be transferred to other European jurisdictions, as the guidelines for heart failure management are consistent across Europe (12) and the determinants for the better outcomes when using remote patient monitoring systems should not change dramatically across different European countries. Still, caution

should be taken for the different organisation of healthcare systems and how that could change the HF patient pathway within the care system. Context can really matter.

For the US, the issue of the healthcare system organisation differences should be further emphasised, as the larger differences in healthcare organisation when compared to European systems can lead to different outcomes when using remote patient monitoring systems. The difference in relative prices of healthcare services – much higher relative savings if you prevent a hospital admission – can also tip the scale in favour of remote monitoring solutions when they reduce hospitalisations. In the opposite direction, there is a cultural difference when compared to European countries, as there is a tendency to avoid legal claims at any price in the US. For that reason, clinicians or healthcare providers might not be willing to take risks when keeping patients at home, particularly for the well-insured, resulting in lower rates of uptake for remote patient monitoring systems.

Patients

There is a dimension of patient preferences that has not been assessed in this thesis. Moving from a face-to-face type of care to a remote environment implies a change in the behaviour of patients and their interaction with the healthcare system, which should be assessed more carefully. Therefore, there is room for co-creating HTM systems together with the patients, based on their needs, preferences, and capabilities.

The COVID-19 pandemic can be used as an example of the success of the transition from the traditional health care facilities to home, as during this time citizens were forced to change their interaction with almost every service to a remote environment. However, it should be noted that heart failure patients tend to be older and, for older age groups, the transition has not been as smooth as for younger age groups. Many actions taken by older citizens in a remote environment normally have the help of younger family members or dedicated carers. So, it should be properly assessed whether changing the reality of heart failure management to a remote environment could not have undesired outcomes for the patients and whether it would be their preference to do so. In addition, although there was an apparent success in the transition to remote care, the real impact of this transition is yet to be properly assessed in terms of health outcomes.

Despite the considerations above, the World does not stop and the transition seems inevitable. It is expected that the use of early warning systems for chronic disease management will become hegemonic in healthcare systems across the globe. Society

should thus be prepared to empower patients and to include them in taking the decisions regarding the transformation that healthcare systems will experience.

Regulators and lawmakers

As previously noted, it is extremely likely that the use of early warning systems – as well as other healthcare interventions supported by digital technologies, eHealth, or mHealth – for chronic disease management will increase in the near future. Thus, the main implications for regulators and lawmakers in the context of this thesis come from the findings presented in **Chapter 7**. It is imperative that the standards for the approval of medical devices, where most of the previously mentioned health technologies fit in, match the ones in force for drugs, especially when it comes to evidence generation on their safety and effectiveness. Those data will be crucial for using health technology assessment methodology for properly appraising new health care interventions leveraged by digital technologies.

There is a need for a conceptual change in the way medical devices are perceived in the healthcare sector, especially when they fall under the umbrella of interventional devices, i.e. health technologies intended for the treatment of a disease, injury or disability. Interventional medical devices ought to be seen as treatments and, as any other treatment in healthcare space, they should be assessed using the best practices of regulatory assessment that is used for drugs, which demands high-quality data. Lawmakers should legislate accordingly, taking into account the risks and uncertainties associated with new health technologies coming into the market and the information needed for correctly deciding upon the incorporation of new health technologies in health care.

Policy makers and payers: reimbursement and clinical guidelines

We have previously touched upon the implications of the findings of the thesis for clinicians and patients, which focused mainly on clinical outcomes. Policy makers and payers are equally interested in the economic outcomes or, in other words, in the cost-effectiveness of the interventions under analysis.

The results presented in **Chapter 6** determine that the home telemonitoring system under analysis, especially when complemented with a diagnostic algorithm, is cost-effective for the management of heart failure in the Dutch setting and, therefore, should be reimbursed and included in the clinical guidelines for heart failure management. It is unwise to extrapolate this recommendation to other countries, as specific costs for each of the resources used may differ, as well as the perspective of the analysis used by the health technology assessment bodies responsible for issuing recommendations on reimbursement of health technologies in those jurisdictions.

On a larger scale, it is critical to raise awareness that the arsenal for providing care is becoming more diverse and that methodology for properly assessing new health technologies should follow that trend. In fact, the Federal Institute for Drugs and Medical Devices in Germany is assessing digital health applications for reimbursement (323). Other countries' policy makers ought to learn from this experience and collaboratively work on solutions for the assessment of healthcare interventions supported by digital technologies, eHealth, and mHealth. Only then can healthcare systems move in the direction of optimal allocation of resources in a value for money framework in the present reality of their existence.

Concluding remarks and recommendations for future research

This thesis contributed to the development of the methodology used in the economic evaluation of early warning systems for chronic disease management. Firstly, by exploring the main concepts embodied in early warning and thus suggesting an overarching definition of early warning systems in the context of health care. Following that definition, methods used in the search of bibliographic electronic databases have also been advanced, mainly through defining sensitive filters that successfully captured studies about early warning systems in the literature review. Although the impact of hospitalisation on health-related quality of life in patients with chronic heart failure was estimated in this thesis, some uncertainty remains in regards to its actual impact, as the data used in the TEN-HMS study were not generated for answering that particular research question. This thesis also presented an ensemble algorithm to predict clinical deterioration in heart failure patients using a telehealth programme, thereby contributing to the advance of the methods that utilise data generated by early warning systems with the goal of predicting and avoiding undesirable events in clinical practice.

The largest contribution of this thesis for the methodology used in the economic evaluation of early warning systems for chronic disease management was the development and validation of a discrete event simulation model able to model heart failure patients and to account for the impact of individual patient characteristics in their health outcomes. The main goal of the thesis was to develop a generic model for assessing the cost-effectiveness of early warning systems. Although the developed model was specific for heart failure patients, the model itself and the discussion around it certainly constitute a stepping stone for the wider goal. The model framework and the open environment in which the model was developed constitute important features that help to recognise that significant contributions were made. Those contributions were tested by estimating the cost-effectiveness analysis of a

home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands. During this endeavour, some challenges regarding available data were encountered, which spurred the review and comparison of the legal framework for the approval of medical devices and drugs and the investigation of the available information on manufacturer sponsored clinical studies and HTA-supported recommendations for medical devices and drugs. From this study and from the overall work done in the thesis, some recommendation for future research can be issued.

Firstly, early warning systems could be introduced as a thesaurus term in the bibliographical databases containing scientific literature. From our point of view, it would be worthwhile to harmonise the definition and to direct efforts of healthcare researchers to find common ground in early warning systems, regardless of the specificities of each intervention falling under that umbrella. This would save time and resources, as some inefficient and repeated work could be avoided.

Secondly, the impact of hospitalisation on health-related quality of life of heart failure patients should be further investigated, as there is limited information on this topic. This is likely to be true for other chronic diseases. Considering that health-related quality of life is such an important factor in decision making in health care, it is critical to have the best possible evidence so that decision can be made with a high degree of confidence.

Thirdly, the developed model could and should be further improved in order to be more comprehensive and more easily usable for other diseases and early warning system interventions. These recommendations would results in lesser adaptations than currently needed and in a more user-friendly environment for the non-experienced programmer in R that would allow for generating outcomes by simply changing key model parameters.

Fourthly, the model should be validated with another dataset. Unfortunately we did not have access to one, but other researchers interested in the topic will be able to do so by accessing the publicly available code and replacing the inputs for the performed analyses.

Fifthly, future economic evaluations studies should be developed in R or other open source software. One problem that we often found was the lack of real access to models, besides the information contained in the publication – which is often oversimplified for publication purposes. In fairness, this seems to be the righteous approach to Science and the one that is more truthful to its core principles.

Sixthly, we recommend that interested parties in other jurisdictions make use of the model and the cost-effectiveness analysis presented in this thesis in order to appropriately transfer results and contribute to the rational decision making on home

telemonitoring systems and diagnostic algorithms in the management of heart failure in their countries.

Seventhly, health economic researchers should liaise with experts in artificial intelligence and machine learning in order to understand how the current developments in those fields can be linked to the topics address in this thesis. From a broad perspective, it seems likely that a coding environment like R can introduce artificial intelligence and machine learning in the developed model, both for the cost-effectiveness assessment and for the diagnostic algorithm predictions. For the algorithm, one can think of constant improvement through checking predictions and outcomes, thus improving the accuracy of the algorithm. As a consequence, this input could be used in the cost-effectiveness calculations. In this way, the improvement of the algorithm predictions and consequent follow-up actions – based on the risk score and the individual patient characteristics – could be driven by the improvement of the overall cost-effectiveness of the intervention.

Finally, besides the implications for different stakeholders that have been previous pointed out, we would like to emphasise that a different approach to early warning systems (also in their capacity of medical devices) is needed. These health technologies will inevitably flood the health care space in the upcoming years. For that reason, it is critical that researchers act preventively and that some serious work is done for developing methods used for accurately assess early warning systems. As the innovation in health care keeps on providing new solutions for patients, the ones who are responsible for assessing and deciding upon their uptake have a moral duty to do it in the best possible way. For we will all be patients one day.

Summary

Background

Chronic diseases represent 71% of worldwide mortality, with an estimated 41 million deaths in 2016. Cardiovascular diseases, cancer, chronic respiratory diseases are responsible for approximately 80% of the death toll of chronic diseases. Along with mortality comes a huge financial burden on health care systems around the world, especially in highly developed countries, where between 70-90% of the money spent on health care is allocated to chronic disease management. The *2030 Agenda for Sustainable Development* adopted by the United Nations recognised chronic disease management as one of the major challenges for improving health of the populations. The World Health Organisation *Sustainable Development Goals* targets by 2030 clearly state that countries need comprehensive strategies to reduce death from chronic diseases more effectively.

Chronic disease management consists of detecting, diagnosing, treating, and monitoring chronic diseases and providing access to palliative care for those in need. Managing chronic diseases is a major challenge for healthcare systems worldwide, which have been primarily designed to address acute episodic care rather than to provide organised care for people with long-term medical conditions. Orthodox chronic disease management models face notorious problems in their implementation, mainly related to the communication between professionals from different disciplines and organisations, the engagement of patients in the self-management of their disease, the recording and keeping track of clinical records, and the burden of regular review and follow-up of patients. Digital technologies (electronic processes and communications, the internet, and other information technologies) – usually labelled under eHealth, or mHealth if involving mobile devices – are seen as part of the solution to these problems. The promise of technology in the treatment of chronic disease rests on two essential pillars: (i) providing a framework for patient engagement in changing modifiable behavioural risk factors and (ii) generating, collecting, processing, and analysing disease-related data that can be used for predicting important events related to the disease and for fine-tuning the treatment of patients.

Early warning systems are timely surveillance systems that collect information on diseases in order to anticipate health deterioration and to trigger prompt clinical intervention, thereby improving prognosis and treatment outcomes. Generally speaking, early warning systems in health care consist of three main elements: (i) monitoring and collection of clinical data (e.g., vital signs, biomarkers, self-reported health status); (ii) a framework allowing for the identification of patterns and trends in these data, indicating significant changes in the health status of the patients; and (iii) the establishment of pre-determined conditions – such as the existence of

statistically uncommon patterns in the data, threshold values or ranges for specific parameters within the collected data, or the presence of a particular combination of signs and symptoms – that trigger an alarm and follow-up actions.

Health technology assessment is the systematic evaluation of properties, effects, and/or impacts of health technology through a multidisciplinary process that assesses the social, economic, organisational and ethical issues of a health intervention or health technology with the main purpose informing a policy decision making. Economic evaluation is a topic of growing interest in the context of the assessment of health technologies, as policy makers have turned to evidence based decision making for supporting their reimbursement decisions. By focusing in the comparison of two or more alternative healthcare interventions in terms of their costs and effects, economic evaluations provide cost-effectiveness results on the implementation of new health technologies versus current standard of care, and they allow for a rational allocation of healthcare resources within a *value for money* framework. Because estimating the cost-effectiveness of an intervention in the health care field inevitably comprises the synthesis of information, the increasing use of economic evaluations for decision making in health care led to higher requirements in terms of analytic methodology, which helped to establish decision models as supporting tools in health technology assessment research. In brief, decision models allow to: (i) synthesise all relevant information in an analytical framework that reflects the possible prognoses and the disease pathways and their relationship with the interventions under evaluation; (ii) consider of all relevant comparators, expanding from randomised control trials, which are normally limited to head-to-head comparisons; (iii) use the appropriate time horizon for the context of decision making by extrapolating both costs and effects into the future; and (iv) address variability and uncertainty in a systematic and/or probabilistic manner.

Objective

The main objective of this thesis is to study the methodology used in the economic evaluation of early warning systems for chronic disease management, with a particular focus on the decision modelling methods used in this framework. Considering that early warning systems, regardless of their target disease, have in common that they are aimed at monitoring patients' health status through periodically measuring individual patient characteristics in order to anticipate health deterioration and to trigger prompt clinical intervention, the possibility to develop a more generic decision model for assessing the cost-effectiveness of early warning systems was explored.

Findings

Chapter 2 consists of a systematic literature review describing the general characteristics of decision models used in the economic evaluation of early warning systems for the management of chronic heart failure patients and the assessment of their methodological quality using the checklist for the critical appraisal of decision-analytic models for health technology assessment developed by Philips et al. (2004). Nine electronic databases were searched, retrieving, after deduplication, 4765 references. From the 27 studies identified for full-text reading, seven studies containing decision models were considered for data extraction and data analysis. The retrieved models showed some variability with regards to their general study characteristics and they displayed satisfactory methodological quality overall. However, the consideration and discussion of any competing theories regarding model structure and disease progression, identification of key parameters and the use of expert opinion, and uncertainty analyses were identified as key areas for improvement in the development of future decision models.

Knowing the impact of hospitalisation on health-related quality of life is particularly relevant for informing cost-effectiveness models designed to assess health technologies aimed at reducing hospital admissions. The estimates for (dis)utility are normally used in the calculation of the effectiveness of interventions, especially when discrete event simulations are the employed modelling technique and hospitalisation is a key event considered in the model. **Chapter 3** aims at determining the impact of nonfatal hospitalisations on the health-related quality of life for a cohort of patients previously diagnosed with heart failure by using their quality of life measurements performed with EQ-5D-3L before and after their hospitalisation. The mean difference between health-related quality of life measurement pre and post hospitalisation was found to be 0.020 [95% CI: -0.020, 0.059] when measured with the EQ-5D index, while a mean decrease of -0.012 [95% CI: -0.043, 0.020] in the utility measured with the visual analogue scale was found. Differences in utility variation according to the primary cause for hospitalisation were encountered: hospitalisations due to respiratory/chest infection and ventricular tachycardia showed an improvement in quality of life when considering the index utilities measured before and after admission, while hospital admissions attributed to atrial fibrillation and myocardial infarction showed a negative variation in index utilities measured before and after hospital admission.

Chapter 4 describes a diagnostic algorithm for predicting the clinical deterioration in heart failure patients using a telehealth remote patient monitoring programme. The main objective of the algorithm is to raise an alarm indicating that the patient should be screened considering the risk to be hospitalised. In other words, the alarm should

be raised when the risk score calculated by the algorithm exceeds a predefined threshold. This threshold was estimated using the therapeutic threshold concept, which aims at determining whether to administer or to withhold a treatment given the expected value of either option. Using the average daily costs of heart failure-related hospitalisation of Medicare beneficiaries with heart failure and the expected average weighted length of stay in hospital with and without the telehealth remote patient monitoring programme, it was estimated that a patient should be screened if the probability of being hospitalised calculated by the algorithm is higher than 11.7%.

Considering the limitations of the therapeutic threshold, the added value of the diagnostic algorithm in a clinical setting would be better investigated through a full cost-effectiveness analysis that could properly assess the costs of implementing and using the technology versus the anticipated effectiveness resulting from the early detection of disease deterioration and the subsequent reduction of hospitalisations. **Chapter 5** presents the construction and validation of a discrete event simulation model that is able to model heart failure patients managed with usual care or an early warning system (with or without a diagnostic algorithm) and to account for the impact of individual patient characteristics in their health outcomes. The model was developed using patient-level data from the TEN-HMS study, coded using R, and validated along the lines of the Assessment of the Validation Status of Health-Economic decision models tool. The model includes 20 patient and disease characteristics and generates 8 different outcomes. It showed robustness and validity of generated outcomes when comparing them to other models addressing the same problem and to external data. Therefore, the patient-level simulation model was deemed suitable to be used for simulating a wide range of outcomes for different patient subgroups and treatment scenarios, as well as providing useful information for guiding research and development of new treatment options by showing the hypothetical impact of these interventions on a large number of important heart failure outcomes.

Chapter 6 assesses the cost-effectiveness a home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands. Three interventions were analysed: (i) usual care (patient management plan implemented by the patient's primary care physician), (ii) home telemonitoring, and (iii) home telemonitoring with the addition of a diagnostic algorithm. According to the Dutch guidelines for economic evaluations in healthcare, a societal perspective was adopted, with considered costs including all costs inside the healthcare sector, patient and family, and other sectors, and productivity losses assessed using the friction cost method. Home telemonitoring with the addition of a diagnostic algorithm extendedly dominated home telemonitoring, while it had an incremental cost-effectiveness ratio of €27,712 per quality-adjusted life year against usual care. The estimated cost-

effectiveness threshold for the population under analysis in the Netherlands according to the proportional shortfall method was €80,000 per quality-adjusted life year. At that threshold, the home telemonitoring with the addition of a diagnostic algorithm intervention had a 96.0% probability of being cost-effective versus usual care and it should be adopted in the Netherlands with a high degree of confidence.

Early warning systems are normally medical devices, which traditionally have been less regulated than drugs. As such, there is a generalised idea that the standards of evidence collection for placing medical devices on the market are generally lower than those for drugs, which may ultimately hinder comparison between those health technologies, in particular when they have the same goal and are targeted at the same disease and/or population. With that in mind, **Chapter 7** reviews and compares the legal framework in the United States and the European Union for the approval of medical devices and drugs, with a particular focus on the changes brought by Regulation (EU) 2017/745. It also compares the available information on clinical research and health technology assessment-supported recommendations in each of the considered jurisdictions for the health technologies under analysis. The developed work found different standards for approval of medical devices and drugs on their quality, safety, and performance/efficacy dimensions, as well as substantially lower number of manufacturer-sponsored clinical studies and health technology assessment-supported recommendations for medical devices versus drugs and an indication of lower standards of evidence used in recommendations for medical devices. Therefore, it was concluded that policy changes ought to be implemented in order to promote an integrated evidence-based assessment system for a better allocation of resources in healthcare, namely: a consensual classification of medical devices from an health technology assessment perspective, which could be used as a guide for generating outcomes in clinical investigation, and the adoption of conditional coverage practices including mandatory post-approval evidence development for performing periodic technology assessments.

The general discussion of the thesis is presented in **Chapter 8**, where the main findings are summarised, discussed, and interpreted in the context of the objectives of the thesis. The implications for different stakeholders resulting from the thesis' findings thesis are explored, as well as the challenges and limitations encountered during the research that led to the thesis. Finally, the discussion ends with some concluding remarks and recommendations for future research.

Samenvatting

Achtergrond

Chronische ziekten vertegenwoordigen 71% van de wereldwijde mortaliteit, waarbij in 2016 naar schatting 41 miljoen patiënten zijn overleden. Hart- en vaatziekten, kanker en chronische respiratoire aandoeningen zijn verantwoordelijk voor 80% van het aantal doden door chronische ziekten. Deze mortaliteit brengt een enorme financiële belasting voor gezondheidszorgstelsels overal ter wereld met zich mee, met name in de hoogontwikkelde landen, waar 70-90% van het geld dat aan de gezondheidszorg wordt besteed, wordt toegewezen aan de medische zorg voor chronische ziekten. In de door de Verenigde Naties aangenomen Agenda 2030 voor Duurzame Ontwikkeling wordt medische zorg voor chronische ziekten onderkend als een van de belangrijkste uitdagingen als het gaat om verbetering van de gezondheid van populaties. In de doelstellingen voor 2030 in het kader van de Duurzame Ontwikkelingsdoelstellingen van de Wereldgezondheidsorganisatie wordt duidelijk aangegeven dat landen brede strategieën nodig hebben om overlijden tengevolge van chronische ziekten effectief te verminderen.

Medische zorg voor chronische ziekten bestaat uit opsporing, diagnostiek, behandeling en follow-up van chronische ziekten, en waar nodig uit het bieden van toegang tot palliatieve zorg. Medische zorg voor chronische ziekten is een grote uitdaging voor gezondheidszorgstelsels wereldwijd, die primair zijn ingesteld op het bieden van tijdelijke acute medische zorg en niet op het bieden van georganiseerde medische zorg voor mensen met chronische medische aandoeningen. Bij conventionele modellen voor medische zorg voor chronische ziekten is er sprake van notoire problemen bij de implementering, die hoofdzakelijk verband houden met de communicatie tussen professionals van verschillende disciplines en organisaties, de betrokkenheid van patiënten bij zelfbehandeling van hun aandoening, het vastleggen van en toegang blijven houden tot klinische gegevens en de belasting van regelmatige herevaluatie en follow-up van patiënten. Digitale technologieën (elektronische processen en informatieuitwisseling, het internet en andere informatietechnologieën) – die gewoonlijk worden aangeduid met e-Health, of m-Health als er mobiele apparaten bij betrokken zijn – worden gezien als onderdeel van de oplossing voor deze problemen. De belofte van technologie bij de behandeling van chronische aandoeningen berust op twee essentiële pijlers: (i) het bieden van wegen voor betrokkenheid van patiënten bij de verandering van voor verandering vatbare uit gedrag voortvloeiende risicofactoren en (ii) het genereren, verzamelen, verwerken en analyseren van ziektegerelateerde gegevens die kunnen worden gebruikt voor het voorspellen van belangrijke voorvallen in verband met de aandoening en voor het nauwkeurig afstemmen van de behandeling van patiënten.

Systemen voor vroegtijdige signalering (early warning systems) zijn systemen voor tijdige controle waarmee gegevens over aandoeningen worden verzameld, met het doel verslechtering van de gezondheidstoestand te zien aankomen en snelle klinische interventie te initiëren, en zo de prognose en behandelingsuitkomsten te verbeteren. In het algemeen gesproken bestaan systemen voor vroegtijdige signalering uit drie hoofdelementen: (i) het detecteren en verzamelen van klinische gegevens (bijv. vitale functies, biomarkers, patiëntgerapporteerde gezondheidstoestand); (ii) een systeem waarmee patronen en trends in deze gegevens kunnen worden herkend, die kunnen wijzen op relevante veranderingen in de gezondheidstoestand van patiënten; en (iii) het vaststellen van gepredetermineerde aandoeningen – zoals aanwezigheid van statistisch ongebruikelijke patronen in de gegevens, drempelwaarden of waardenbereiken voor specifieke parameters binnen de verzamelde gegevens of de aanwezigheid van een bepaalde combinatie van klachten en symptomen – die een waarschuwing en follow-upacties genereren.

Health technology assessment is de systematische evaluatie van eigenschappen, effecten en/of gevolgen van medische technologie door middel van een multidisciplinair proces waarbij ook de sociale, economische, organisatorische en ethische kwesties van een medische interventie of medische technologie worden beoordeeld, met als hoofddoel materiaal aan te bieden voor besluitvorming rond een beleid. Economische evaluatie is een onderwerp dat in de context van beoordeling van medische technologieën toenemende belangstelling geniet, omdat beleidsvormers zich zijn gaan toeleggen op evidence-based besluitvorming ter ondersteuning van hun vergoedingsbeslissingen. Economische evaluaties – waarbij twee of meer medische interventies wat betreft kosten en effecten met elkaar worden vergeleken – voorzien in kosteneffectiviteitsresultaten met betrekking tot de implementering van nieuwe medische technologieën in vergelijking met de standaard medische zorg, en ze maken een rationele toewijzing van gezondheidszorgmiddelen binnen een 'value for money'-kader mogelijk. Omdat bij de bepaling van de kosteneffectiviteit van een interventie op het gebied van de gezondheidszorg onvermijdelijk informatie moet worden samengesteld, heeft het toenemend gebruik van economische evaluaties voor besluitvorming in de gezondheidszorg geleid tot hogere eisen wat betreft analytische methodologie, zodat er beslissingsmodellen zijn ingesteld als ondersteunende middelen bij health technology assessment research. Om kort te gaan, beslissingsmodellen bieden de mogelijkheid: (i) alle relevante informatie samen te stellen binnen een analytische structuur die een overzicht biedt van de mogelijke prognoses en ziekteverlopen en hun relatie met de te evalueren interventies; (ii) alle relevante comparatoren in aanmerking te nemen, in aanvulling op gerandomiseerde gecontroleerde onderzoeken, die zich gewoonlijk beperken tot rechtstreekse vergelijkingen; (iii) voor de context van de besluitvorming de gepaste tijdshorizon te gebruiken door zowel kosten als effecten naar de toekomst te extrapoleren; en (iv)

variabiliteit en onzekerheid op een systematische en/of probabilistische wijze te behandelen.

Doel

Het hoofddoel van deze thesis is onderzoek naar de methodologie die wordt gebruikt bij de economische evaluatie van systemen voor vroegtijdige signalering in de medische zorg voor chronische ziekten, met speciale aandacht voor de methoden voor beslissingsmodellen die in dit kader worden gebruikt. Gezien dat systemen voor vroegtijdige signalering, ongeacht de aandoening waarvoor ze worden gebruikt, gemeenschappelijk hebben dat ze zich richten op bewaking van de gezondheidstoestand van de patiënt via periodieke bepaling van individuele kenmerken van de patiënt – teneinde verslechtering van de gezondheidstoestand te zien aankomen en snelle klinische interventie te initiëren – is onderzoek gedaan naar de mogelijkheid om een algemener beslissingsmodel voor beoordeling van de kosteneffectiviteit van systemen voor vroegtijdige signalering te ontwikkelen.

Bevindingen

Hoofdstuk 2 bestaat uit een systematische literatuurstudie met beschrijving van de algemene kenmerken van beslissingsmodellen die worden gebruikt bij de economische evaluatie van systemen voor vroegtijdige signalering in de medische zorg voor patiënten met chronisch hartfalen en beoordeling van de methodologische kwaliteit van deze modellen met de door Philips et al. ontwikkelde checklist (2004) voor kritische waardering van besliskundige modellen voor health technology assessment. Er zijn negen elektronische databases doorzocht, waarbij, na deduplicatie, 4765 referenties zijn verkregen. Van de 27 studies die zijn gevonden voor lezen van de volledige tekst, zijn zeven studies met beslissingsmodellen in aanmerking genomen voor gegevensextractie en gegevensanalyse. Bij de verkregen modellen was er sprake van enige variabiliteit met betrekking tot de algemene kenmerken van de corresponderende studies en de modellen hadden over het geheel genomen een toereikende methodologische kwaliteit. Er zijn echter belangrijke elementen voor verbetering van de ontwikkeling van toekomstige beslissingsmodellen geïdentificeerd, zoals bestudering en bespreking van concurrerende theorieën met betrekking tot modelstructuur en ziekteprogressie, identificatie van belangrijke parameters en het gebruik van deskundigenadvies en onzekerheidsanalyses.

Kennis van de gevolgen van ziekenhuisopname op gezondheidsgerelateerde kwaliteit van leven is met name van belang bij het opstellen van kosteneffectiviteitsmodellen die zijn bestemd voor de evaluatie van medische technologie die is gericht op vermindering van het aantal ziekenhuisopnames. De schattingen voor (dis)utiliteit

worden gewoonlijk gebruikt in de berekening van de effectiviteit van interventies, met name wanneer 'discrete event'-simulaties als techniek voor het opstellen van modellen worden gebruikt en ziekenhuisopname als een belangrijk voorval in het model wordt beschouwd.

Hoofdstuk 3 richt zich op bepaling van het effect van ziekenhuisopnames zonder fatale afloop op de gezondheidsgerelateerde kwaliteit van leven voor een groep patiënten bij wie eerder de diagnose hartfalen is gesteld, door gebruik te maken van de metingen van hun kwaliteit van leven die voorafgaand aan en na ziekenhuisopname met de EQ-5D-3L-vragenlijst waren uitgevoerd. Het gemiddelde verschil tussen de meting van gezondheidsgerelateerde kwaliteit van leven voorafgaand aan en na de ziekenhuisopname bleek 0,020 [95%-BI: -0,020, 0,059] te zijn bij meting met de EQ-5D-3L-vragenlijst, terwijl de utiliteit die met de visueel analoge schaal was gemeten een gemiddelde afname van -0,012 [95%-BI: -0,043, 0,020] liet zien. Er werden verschillen in utiliteiten aangetroffen volgens de primaire reden voor ziekenhuisopname: bij ziekenhuisopnames vanwege een onderste luchtweginfectie en ventriculaire tachycardie werd er een verbetering van de kwaliteit van leven gezien op basis van de utiliteiten die voorafgaand aan en na de ziekenhuisopname met de EQ-5D-3L-vragenlijst waren gemeten, terwijl bij ziekenhuisopnames vanwege atriumfibrilleren en myocardinfarct op basis van de utiliteiten bij dezelfde meting een vermindering van de kwaliteit van leven werd gezien.

Hoofdstuk 4 beschrijft een diagnostisch algoritme voor voorspelling van klinische achteruitgang bij patiënten met hartfalen bij gebruik van een telegeneeskundig programma voor patiëntbewaking op afstand. Het voornaamste doel van het algoritme is om een waarschuwing te genereren wanneer een patiënt dient te worden gecontroleerd gezien het risico van ziekenhuisopname. Met andere woorden, de waarschuwing dient te worden gegenereerd wanneer de risicoscore die door het algoritme wordt berekend boven een vooraf gedefinieerde drempelwaarde komt. De drempelwaarde werd berekend door gebruik te maken van het concept behandeldrempel, waarbij het gaat om beantwoording van de vraag een behandeling geven of geen behandeling geven op basis van de verwachte afzonderlijke waarde van beide opties. Bij gebruik van de gemiddelde dagelijkse kosten van ziekenhuisopname in verband met hartfalen, in het geval van patiënten met hartfalen die in aanmerking komen voor Medicare-vergoedingen, en de verwachte gewogen gemiddelde duur van verblijf in het ziekenhuis met en zonder gebruik van het telegeneeskundige programma voor patiëntbewaking op afstand, werd er berekend dat een patiënt dient te worden gecontroleerd als de waarschijnlijkheid van ziekenhuisopname volgens berekening met het algoritme hoger is dan 11,7%.

Gezien de beperkingen van de drempelwaarde voor behandeling, zou het beter zijn om de toegevoegde waarde van het diagnostische algoritme in een klinische setting te onderzoeken door middel van een volledige kosteneffectiviteitsanalyse, waarbij de kosten van de invoering en het gebruik van de technologie naar behoren kunnen worden bepaald in vergelijking met de verwachte effectiviteit als gevolg van de vroegtijdige detectie van klinische achteruitgang en de daaropvolgende vermindering van het aantal ziekenhuisopnames.

Hoofdstuk 5 presenteert de samenstelling en validatie van een 'discrete event'-simulatiemodel dat kan dienen als een model voor patiënten met hartfalen die de gebruikelijke medische zorg krijgen of bij wie een systeem voor vroegtijdige signalering (met of zonder een diagnostisch algoritme) wordt gebruikt en waarmee het effect van individuele kenmerken van de patiënt op diens gezondheidsuitkomsten kan worden verklaard. Het model, gecodeerd met R, is ontwikkeld met gegevens van individuele patiënten uit het TEN-HMS-onderzoek en gevalideerd volgens de principes van het hulpmiddel Assessment of the Validation Status of Health-Economic decision models. In het model zijn 20 patiënt- en ziektekenmerken opgenomen en het model genereert 8 verschillende uitkomsten. De uitkomsten die het model genereert, zijn robuust en geldig bij vergelijking van de uitkomsten met die van andere modellen voor hetzelfde probleem en bij vergelijking van de uitkomsten met externe gegevens. Het simulatiemodel voor individuele patiënten werd daarom geschikt geacht om te worden gebruikt voor simulatie van een breed scala van uitkomsten voor verschillende patiëntensubgroepen en behandelingsscenario's, en ook om te voorzien in informatie die kan worden gebruikt om richting te geven aan onderzoek en ontwikkeling van nieuwe behandelingsopties, door het hypothetische effect van deze interventies op een groot aantal belangrijke uitkomsten bij hartfalen aan te tonen.

Hoofdstuk 6 is een hoofdstuk waarin de kosteneffectiviteit van een systeem voor telemonitoring thuis en een diagnostisch algoritme worden beoordeeld in de context van medische zorg voor hartfalen in Nederland. Er zijn drie interventies geanalyseerd: (i) gebruikelijke medische zorg (zorgplan voor de patiënt dat door de eerstelijnsarts van de patiënt wordt geïmplementeerd), (ii) telemonitoring thuis en (iii) telemonitoring thuis met toevoeging van een diagnostisch algoritme. Volgens de Nederlandse richtlijnen voor economische evaluaties in de gezondheidszorg is er een maatschappelijk perspectief gehanteerd, waarbij kosten in aanmerking worden genomen zoals alle kosten binnen de gezondheidszorgsector, kosten voor de patiënt en het gezin, kosten binnen andere sectoren en de kosten van productiviteitsverlies die met de frictiekostenmethode worden bepaald. Telemonitoring thuis werd indirect gedomineerd door telemonitoring thuis met de toevoeging van een diagnostisch algoritme, terwijl deze combinatie ten opzichte van de gebruikelijke medische zorg een incrementele kosteneffectiviteitsratio van € 27.712 per kwaliteit van leven

gecorrigeerd levensjaar had. De berekende drempelwaarde voor kosteneffectiviteit met betrekking tot de te analyseren populatie in Nederland was volgens de 'proportional shortfall'-methode € 80.000 per kwaliteit van leven gecorrigeerd levensjaar. Bij die drempelwaarde gold voor telemonitoring thuis met toevoeging van een diagnostisch algoritme een waarschijnlijkheid van 96% dat deze interventie kosteneffectief was in vergelijking met gebruikelijke medische zorg, en deze interventie zou in Nederland met een hoge graad van zekerheid moeten worden aanvaard.

Systemen voor vroegtijdige signalering zijn meestal medische hulpmiddelen, waarop in het verleden minder wet- en regelgeving van toepassing is geweest dan op geneesmiddelen. Als zodanig is er een wijdverbreid idee dat de normen voor verzameling van gegevens op grond waarvan medische hulpmiddelen in de handel kunnen worden gebracht, in het algemeen lager zijn dan die voor geneesmiddelen, wat uiteindelijk een belemmering kan vormen voor de vergelijking van deze medische technologieën, met name wanneer ze hetzelfde doel hebben en op dezelfde aandoening en/of populatie zijn gericht.

Met het oog daarop wordt in **hoofdstuk 7** de wet- en regelgeving voor de goedkeuring van medische hulpmiddelen en geneesmiddelen in de Verenigde Staten en de Europese Unie besproken en vergeleken, waarbij met name aandacht wordt besteed aan de veranderingen die Verordening (EU) 2017/745 met zich heeft meegebracht. In het hoofdstuk worden in de betreffende rechtsgebieden met betrekking tot de te analyseren medische technologieën ook de beschikbare informatie over klinisch-wetenschappelijk onderzoek en door health technology assessment ondersteunde aanbevelingen vergeleken. Op grond van het werk dat in deze richting is verricht, is vastgesteld dat voor medische hulpmiddelen en geneesmiddelen wat betreft de kenmerken kwaliteit, veiligheid en werking/werkzaamheid verschillende normen voor goedkeuring gelden, dat het aantal in opdracht van fabrikanten uitgevoerde klinische onderzoeken en door health technology assessment ondersteunde aanbevelingen aanzienlijk lager is voor medische hulpmiddelen dan voor geneesmiddelen en dat er lagere normen lijken te worden gehanteerd voor gegevens die aanbevelingen voor medische hulpmiddelen ondersteunen. Daarom werd geconcludeerd dat er beleidsveranderingen dienen te worden doorgevoerd ter bevordering van een geïntegreerd evidence-based beoordelingssysteem voor een betere toewijzing van middelen in de gezondheidszorg, namelijk: een op consensus berustende classificatie van medische hulpmiddelen vanuit het gezichtspunt van health technology assessment, die gebruikt zou kunnen worden als richtsnoer voor het genereren van uitkomsten in klinisch onderzoek en invoering van regels voor voorwaardelijke dekking, inclusief verplichte postmarketing verzameling van relevante gegevens voor uitvoering van periodieke technology assessments.

De algemene bespreking van deze thesis wordt gepresenteerd in **hoofdstuk 8**, waarin de belangrijkste bevindingen worden samengevat, besproken en geïnterpreteerd in de context van de doelstellingen van de thesis. Er wordt op grond van de bevindingen van de thesis gekeken naar de implicaties voor de verschillende stakeholders en ook naar de moeilijkheden en beperkingen die werden ervaren tijdens het onderzoek dat tot de thesis heeft geleid. De bespreking eindigt met enkele afsluitende opmerkingen en aanbevelingen voor toekomstig onderzoek.

PhD portfolio

PhD candidate	Fernando Guilherme França Pereira Albuquerque de Almeida
Doctoral supervisor	prof. dr. M.P.M.H. Rutten-van Mölken
Daily advisor	dr. M.J. Al dr. I. Corro Ramos
PhD period	October 2014 – July 2021

	Year	ECTS
Academic training		
Courses & workshops		
Introduction to Statistics Descriptive Statistics (edX – Berkeley X)	2014	1.5
Quantitative methods in clinical & Public Health research (edX – HarvardX)	2014	2.0
How to use EndNote (Erasmus MC)	2014	0.3
Systematic literature retrieval in PubMed, part I (Erasmus MC)	2014	0.3
Systematic literature retrieval in PubMed, part II (Erasmus MC)	2014	0.3
Systematic literature retrieval in other databases (Erasmus MC)	2014	0.3
Academic Integrity Day (Master Class/EUR)	2015	0.5
English academic writing for PhD students (EUR)	2015	0.3
Biostatistics for clinicians (NIHES/Erasmus MC)	2015	1.0
Regression analysis for clinicians (NIHES/Erasmus MC)	2015	1.0
Survival analysis for clinicians (NIHES/Erasmus MC)	2015	1.0
Project management – Klaar in 4 jaar (EUR)	2015	0.2
Decision analytic modelling for economic evaluation (The University of York)	2015	1.5
R Programming (Coursera – John Hopkins University)	2016	2.0
Writing in Sciences (Coursera – Stanford University)	2016	1.0
Patient-level simulation modelling in R (ESHPM/EUR)	2017	1.0
Network meta-analysis (EUR)	2018	0.5

Symposia & congresses		
ISPOR 17th Annual European Congress	2015	1.0
lolaHESG	2017	1.0
Virtual ISPOR Europe 2020	2020	1.0
Other		
Dutch courses for EUR-employees	2015-2017	3.0
Reviewer for scientific publication (1)	2016	0.2
Member of the ERG team in a project for NICE	2016	0.8
President of the ESHPM ISPOR Student Chapter	2017-2018	1.5
Teaching activities		
Mentoring, tutoring, and lecturing		
Markov modelling (post-graduation/Universidade Católica Portuguesa)	2015	0.5
Health Technology Assessment (exam correction)	2014-2015	1.0
Health Technology Assessment (Master)	2014-2018	4.0
Pharmaceutical Pricing and Market Access (Master)	2014-2018	4.0
Participating in HTA Research (Master)	2014-2016	2.0
Advanced Health Economic Modelling (Master)	2016-2018	2.0
Supervision		
Master thesis (1 student)	2014-2015	1.5
Master thesis (1 student)	2015-2016	1.5
Master thesis (3 students)	2016-2017	4.5
Master thesis (3 students)	2017-2018	4.5
TOTAL		**48.7**

List of publications

Included in this dissertation

Albuquerque de Almeida, F., Ricardo, M. (2021). Medical devices rise: "We demand the same rights!". Health Policy and Technology (Submitted).

Albuquerque de Almeida, F., Corro Ramos, I., Rutten-van Mölken, M., & Al, M. (2021). Cost-effectiveness of a home telemonitoring system and a diagnostic algorithm in the management of heart failure in the Netherlands. JMIR mHealth and uHealth (In review).

Albuquerque de Almeida, F., Corro Ramos, I., Rutten-van Mölken, M., & Al, M. (2021). Modeling Early Warning Systems: Construction and Validation of a Discrete Event Simulation Model for Heart Failure. Value in Health (In Press).

https://doi.org/10.1016/j.jval.2021.04.004

Albuquerque de Almeida, F., Al, M.J., Koymans, R., Riistama, J., Pauws, S., & Severens, J.L. (2020). Impact of hospitalisation on health-related quality of life in patients with chronic heart failure. Health and Quality of Life Outcomes 18(1): 1-10.

Albuquerque de Almeida, F., Al, M., Koymans, R., Caliskan, K., Kerstens, A., & Severens, J. L. (2018). Early warning systems for the management of chronic heart failure: a systematic literature review of cost-effectiveness models. Expert review of pharmacoeconomics & outcomes research, 18(2), 161-175.

Not included in this dissertation

Albuquerque de Almeida, F. (2016). Introdução à Avaliação Económica de Tecnologias de Saúde. Lisboa: Pharmavalue | Focus on Evolution.

Büyükkaramikli, N. C., de Groot, S., Fayter, D., Wolff, R., Armstrong, N., Stirk, L., Worthy, G., **Albuquerque de Almeida, F.**, Kleijnen, J., & Al, M. J. (2018). Pomalidomide with dexamethasone for treating relapsed and refractory multiple myeloma previously treated with lenalidomide and bortezomib: an evidence review group perspective of an NICE single technology appraisal. PharmacoEconomics, 1-15.

Portela, M. D. C. C., Sinogas, C., **Albuquerque de Almeida, F.**, Baptista-Leite, R., & Castro-Caldas, A. (2017). Biologicals and biosimilars: safety issues in Europe. Expert opinion on biological therapy, 17(7), 871-877.

Portela, M. D. C. C., Sinogas, C., **Albuquerque de Almeida, F.**, Baptista-Leite, R., & Castro-Caldas, A. (2017). Biologicals and Biosimilars: Gaps in the Pharmacovigilance System in Portugal. Acta medica portuguesa, 30(3), 205-212.

Acknowledgements

To everyone who stood by me during this journey, for you were the ones who suffered the most. I could not have written this thesis without your unwavering support.

About the author

Fernando Albuquerque de Almeida was born on 17 August 1986 in Lisbon. After following the scientific track in high school, he joined the University of Lisbon in 2005 for pursuing his education in Pharmaceutical Sciences. During his track, Fernando did his ERASMUS research programme in the Freie Universität Berlin, where he developed the thesis entitled *Hot-melt extrusion: Improving the solubility of carbamazepine* with which he successfully defended his PharmD. For his academic record during the PharmD, Fernando received a "Merit Scholarship" awarded by the University of Lisbon.

In early 2012, he started working as a quality assurance and quality control officer in OM Pharma. After obtaining a grant for postgraduate studies in Health Economics awarded by the Merck Sharp & Dohme Foundation during that same year, Fernando moved to the Netherlands to pursue his MSc in Health Economics, Policy and Law (specialization in Health Technology Assessment) at the Erasmus School of Health Policy and Management (ESHPM) in Erasmus University Rotterdam. He concluded his masters one year later, defending his dissertation titled *Transferability assessment of a NICE single technology appraisal (STA) to Portugal – the case of cabazitaxel for the second-line treatment of metastatic hormone refractory prostate cancer*, receiving a "Certificate of Honour" from ESHPM in the process. Between 2013 and 2014, he worked remotely from Lisbon as a consultant in the fields of Pricing and Reimbursement, Public Affairs, and Market Access.

In October 2014, Fernando moved back to the Netherlands to start his PhD, joining a joint research project between Philips Research and the Erasmus School of Health Policy and Management in the field of decision analytic modelling for the economic evaluation of chronic disease management programmes. In late 2018, he returned to Portugal, where he continued working as an external PhD candidate at ESHPM. In March 2019, he started working in Pfizer as a Pricing and Access Manager. In March 2021, Fernando started a new role as a Director in Health Economics and Outcomes Research in a global team within the company.

References

1. Centers for Disease Control and Prevention (CDC). National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). About Chronic Diseases 2021 [12/03/2021]. Available from: https://www.cdc.gov/chronicdisease/about/index.htm.

2. World Health Organization. Global health estimates 2016: deaths by cause, age, sex, by country and by region, 2000-2016. 2018 [12/03/2021]. Available from: https://www.who.int/healthinfo/global_burden_disease/estimates/en/.

3. World Health Organization (WHO). Health technology assessment [25/01/2021]. Available from: https://www.who.int/teams/health-product-and-policy-standards/access-to-assistive-technology-medical-devices/medical-devices/assessment.

4. Buttorff C, Ruder T, Bauman M. Multiple chronic conditions in the United States: Rand Santa Monica, CA; 2017.

5. Martin AB, Hartman M, Lassman D, Catlin A, Team NHEA. National Health Care Spending In 2019: Steady Growth For The Fourth Consecutive Year: Study examines national health care spending for 2019. Health Affairs. 2021:10.1377/hlthaff. 2020.02022.

6. European Commision. Non-communicable diseases [12/03/2021]. Available from: https://ec.europa.eu/health/non_communicable_diseases/overview_en.

7. Directorate for Health and Food Safety EC. Towards better prevention and management of chronic diseases 2016 [12/03/2021]. Available from: https://ec.europa.eu/health/newsletter/169/focus_newsletter_en.htm.

8. World Health Organization. World health statistics 2020: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2020 2020.

9. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-22.

10. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Journal of the American College of Cardiology. 2013;62(16):e147-e239.

11. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Journal of the American College of Cardiology. 2017;70(6):776-803.

12. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. European heart journal. 2016;37(27):2129-200.

13. Butler J, Fonarow GC, Zile MR, Lam CS, Roessig L, Schelbert EB, et al. Developing therapies for heart failure with preserved ejection fraction: current state and future directions. JACC: Heart Failure. 2014;2(2):97-112.

14. Levin R, Dolgin M, Fox C, Gorlin R. The Criteria Committee of the New York Heart Association: Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. LWW Handbooks. 1994;9:344.

15. Pieske B, Tschöpe C, De Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA–PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). European heart journal. 2019;40(40):3297-317.

16. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart. 2007;93(9):1137-46.

17. Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW, et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure: The Rotterdam Study. European heart journal. 2004;25(18):1614-9.

18. Dunlay SM, Roger VL. Understanding the epidemic of heart failure: past, present, and future. Current heart failure reports. 2014;11(4):404-15.

19. Roth GA, Forouzanfar MH, Moran AE, Barber R, Nguyen G, Feigin VL, et al. Demographic and epidemiologic drivers of global cardiovascular mortality. New England Journal of Medicine. 2015;372(14):1333-41.

20. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. International journal of cardiology. 2014;171(3):368-76.

21. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation. 2019;139(10):e56-e528.

22. American Heart Association. Classes of Heart Failure 2017 [17/03/2021]. Available from: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure.

23. Hobbs F, Kenkre J, Roalfe A, Davis R, Hare R, Davies M. Impact of heart failure and left ventricular systolic dysfunction on quality of life. A cross-sectional study comparing common chronic cardiac and medical disorders and a representative adult population. European heart journal. 2002;23(23):1867-76.

24. Zambroski CH, Moser DK, Bhat G, Ziegler C. Impact of symptom prevalence and symptom burden on quality of life in patients with heart failure. European Journal of Cardiovascular Nursing. 2005;4(3):198-206.

25. Bekelman DB, Havranek EP, Becker DM, Kutner JS, Peterson PN, Wittstein IS, et al. Symptoms, depression, and quality of life in patients with heart failure. Journal of cardiac failure. 2007;13(8):643-8.

26. Lewis EF, Lamas GA, O'Meara E, Granger CB, Dunlap ME, McKelvie RS, et al. Characterization of health-related quality of life in heart failure patients with preserved versus low ejection fraction in CHARM. European journal of heart failure. 2007;9(1):83-91.

27. Juenger J, Schellberg D, Kraemer S, Haunstetter A, Zugck C, Herzog W, et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. Heart. 2002;87(3):235-41.

28. Alla F, Zannad F, Filippatos G. Epidemiology of acute heart failure syndromes. Heart failure reviews. 2007;12(2):91-5.

29. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. American Heart Association Council on epidemiology and prevention statistics committee and stroke statistics subcommittee. Heart disease and stroke statistics-2018 update: a report from the American Heart Association Circulation. 2018;137(12):e67-e492.

30. Braunwald E. The war against heart failure: the Lancet lecture. The Lancet. 2015;385(9970):812-24.

31. Dunlay SM, Redfield MM, Weston SA, Therneau TM, Hall Long K, Shah ND, et al. Hospitalizations after heart failure diagnosis: a community perspective. Journal of the American College of Cardiology. 2009;54(18):1695-702.

32. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. European journal of heart failure. 2020;22(8):1342-56.

33. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Køber L, Squire IB, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. European heart journal. 2013;34(19):1404-13.

34. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, et al. Risk prediction in patients with heart failure: a systematic review and analysis. JACC: Heart Failure. 2014;2(5):440-6.

35. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC: Heart Failure. 2014;2(5):429-36.

36. Lupón J, De Antonio M, Vila J, Peñafiel J, Galán A, Zamora E, et al. Development of a novel heart failure risk tool: the Barcelona bio-heart failure risk calculator (BCN bio-HF calculator). PloS one. 2014;9(1):e85466.

37. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle heart failure model. Circulation. 2006;113(11):1424-33.

38. Mozaffarian D, Anker SD, Anand I, Linker DT, Sullivan MD, Cleland JG, et al. Prediction of mode of death in heart failure. Circulation. 2007;116(4):392-8.

39. Stewart S, Jenkins A, Buchan S, McGuire A, Capewell S, McMurray JJ. The current cost of heart failure to the National Health Service in the UK. European journal of heart failure. 2002;4(3):361-71.

40. Gheorghiade M, Shah AN, Vaduganathan M, Butler J, Bonow RO, Rosano GM, et al. Recognizing hospitalized heart failure as an entity and developing new therapies to improve outcomes: academics', clinicians', industry's, regulators', and payers' perspectives. Heart failure clinics. 2013;9(3):285-90.

41. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. Journal of the American College of Cardiology. 2014;63(12):1123-33.

42. World Health Organization. Management of noncommunicable diseases [13/03/2021]. Available from: https://www.who.int/activities/management-of-noncommunicable-diseases.

43. Organization WH. Innovative care for chronic conditions: building blocks for actions: global report: World Health Organization; 2002.

44. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Annals of Internal Medicine. 2003;138(3):256-61.

45. Ham C. The Ten Characteristics of the High-Performing Chronic Care Systems. Health Econ Pol'y & L. 2010;5:71.

46. Phillips Jr RL, Starfield B. Why does a US primary care physician workforce crisis matter? American family physician. 2004;70(3):440.

47. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. Jama. 2002;288(14):1775-9.

48. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the chronic care model in the new millennium. Health affairs. 2009;28(1):75-85.

49. Dennis SM, Zwar N, Griffiths R, Roland M, Hasan I, Powell Davies G, et al. Chronic disease management in primary care: from evidence to policy. Medical Journal of Australia. 2008;188:S53-S6.

50. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. New England journal of medicine. 2003;348(26):2635-45.

51. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. Jama. 2007;297(8):831-41.

52. Harris MF, Zwar NA. Care of patients with chronic disease: the challenge for general practice. Medical Journal of Australia. 2007;187(2):104-7.

53. Harris MF, Williams AM, Dennis SM, Zwar NA, Powell Davies G. Chronic disease self-management: implementation with and within Australian general practice. Medical Journal of Australia. 2008;189:S17-S20.

54. Jerant AF, von Friederichs-Fitzwater MM, Moore M. Patients' perceived barriers to active self-management of chronic conditions. Patient education and counseling. 2005;57(3):300-7.

55. Georgeff M. Patients and technology: Digital technologies and chronic disease management. Australian family physician. 2014;43(12):842-6.

56. Milani RV, Lavie CJ. Health care 2020: reengineering health care delivery to combat chronic disease. The American journal of medicine. 2015;128(4):337-43.

57. World Health Organization. eHealth at WHO 2018 [13/03/2021]. Available from: https://www.who.int/ehealth/about/en/.

58. World Health Organization. Emergencies preparedness, response, "Early warning systems", 2015 [31 August 2015]. Available from: http://www.who.int/csr/labepidemiology/projects/ewarn/en/.

59. Albuquerque De Almeida F, Al M, Koymans R, Caliskan K, Kerstens A, Severens JL. Early warning systems for the management of chronic heart failure: a systematic literature review of cost-effectiveness models. Expert review of pharmacoeconomics & outcomes research. 2018;18(2):161-75.

60. Health Do. Comprehensive critical care: a review of adult critical care services. 2000.

61. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database of Systematic Reviews. 2007(3).

62. Field MJ. Telemedicine: A guide to assessing telecommunications for health care. 1996.

63. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. The Lancet. 2011;378(9792):731-9.

64. Suter P, Suter WN, Johnston D. Theory-based telehealth and patient empowerment. Population health management. 2011;14(2):87-92.

65. Tu JV, Jaglal SB, Naylor CD. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. Circulation. 1995;91(3):677-84.

66. Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. Journal of the American College of Cardiology. 2001;37(4):992-7.

67. Laupacis A, Sekar N. Clinical prediction rules: a review and suggested modifications of methodological standards. Jama. 1997;277(6):488-94.

68. Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. Jama. 2003;290(19):2581-7.

69. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Jama. 2000;284(1):79-84.

70. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology. 2017;2(4):230-43.

71. Beam AL, Kohane IS. Big data and machine learning in health care. Jama. 2018;319(13):1317-8.

72. National Information Center on Health Services Research and Health Care Technology (NICHSR). HTA 101: II. FUNDAMENTAL CONCEPTS 2019 [15/03/2021]. Available from: https://www.nlm.nih.gov/nichsr/hta101/ta10104.html.

73. European Medicines Agency. Health technology assessment bodies [15/03/2021]. Available from: https://www.ema.europa.eu/en/partners-networks/health-technology-assessment-bodies.

74. Banta D, Jonsson E. History of HTA: introduction. International journal of technology assessment in health care. 2009;25(S1):1-6.

75. European Network for Health Technology Assessment (EUnetHTA). [23/12/2020]. Available from: https://eunethta.eu/.

76. European Network for Health Technology Assessment (EUnetHTA). EUnetHTA JA2 WP8 DELIVERABLE - HTA Core Model Version 3.0. 2016.

77. Boadway R, Bruce N. Welfare Economics Basil Blackwell. Cambridge; 1984.

78. Hurley J. An Overview of the Normative Economics of the Health Care Sector: in: Culyer. Newhouse; 2000.

79. Brouwer WB, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! Journal of health economics. 2000;19(4):439-59.

80. Brouwer WB, Culyer AJ, van Exel NJA, Rutten FF. Welfarism vs. extra-welfarism. Journal of health economics. 2008;27(2):325-38.

81. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.

82. Jakubiak-Lasocka J, Jakubczyk M. Cost-effectiveness versus cost-utility analyses: what are the motives behind using each and how do their results differ?—A Polish example. Value in health regional issues. 2014;4:66-74.

83. Klarman HE, Rosenthal GD. Cost effectiveness analysis applied to the treatment of chronic renal disease. Medical care. 1968;6(1):48-54.

84. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. Operations research. 1970;18(6):1021-66.

85. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health services research. 1972;7(2):118.

86. Weinstein MC, Russell LB, Gold MR, Siegel JE. Cost-effectiveness in health and medicine: Oxford university press; 1996.

87. Brouwer W, van Baal P, van Exel J, Versteegh M. When is it too expensive? Cost-effectiveness thresholds and health care decision-making. Springer; 2019.

88. Vallejo-Torres L, García-Lorenzo B, Castilla I, Valcárcel-Nazco C, García-Pérez L, Linertová R, et al. On the estimation of the cost-effectiveness threshold: why, what, how? Value in Health. 2016;19(5):558-66.

89. Thokala P, Ochalek J, Leech AA, Tong T. Cost-effectiveness thresholds: the past, the present and the future. Pharmacoeconomics. 2018;36(5):509-22.

90. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Medical care. 1994:150-63.

91. Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios. Health Economics. 1995;4(5):373-81.

92. Van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health economics. 1994;3(5):309-19.

93. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in ecomomic evaluation: an unavoidable fact of life. Health economics. 1997;6(3):217-27.

94. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation: Oup Oxford; 2006.

95. Buxton MJ. Economic evaluation and decision making in the UK. Pharmacoeconomics. 2006;24(11):1133-42.

96. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. 2014;129(3):e28.

97. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States a policy statement from the American Heart Association. Circulation: Heart Failure. 2013;6(3):606-19.

98. Nichols M, Townsend N, Scarborough P, Rayner M. European cardiovascular disease statistics2012.

99. American Heart Association. Classes of Heart Failure 2015 [28 August 2015]. Available from: http://www.heart.org/HEARTORG/Conditions/HeartFailure/AboutHeartFailure/Classes-of-Heart-Failure_UCM_306328_Article.jsp.

100. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. European journal of heart failure. 2012;14(8):803-69.

101. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. Journal of the American College of Cardiology. 2009;54(18):1683-94.

102. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: abridged Cochrane Review. European journal of heart failure. 2011;13(9):1028-40.

103. Mejhert M, Kahan T, Persson H, Edner M. Limited long term effects of a management programme for heart failure. Heart. 2004;90(9):1010-5.

104. Titler MG, Jensen GA, Dochterman JM, Xie XJ, Kanak M, Reed D, et al. Cost of hospital care for older adults with heart failure: medical, pharmaceutical, and nursing costs. Health services research. 2008;43(2):635-55.

105. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

106. Mauskopf J, Paul J, Grant D, Stergachis A. The Role of Cost—Consequence Analysis in Healthcare Decision—Making. PharmacoEconomics. 1998;13(3):277-88.

107. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford; New York: Oxford University Press; 2005.

108. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value in Health. 2003;6(1):9-17.

109. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-9.

110. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. 2004.

111. Burri H, Sticherling C, Wright D, Makino K, Smala A, Tilden D. Cost–consequence analysis of daily continuous remote monitoring of implantable cardiac defibrillator and resynchronization devices in the UK. Europace. 2013;15(11):1601-8.

112. Klersy C, De Silvestri A, Gabutti G, Raisaro A, Curti M, Regoli F, et al. Economic impact of remote patient monitoring: an integrated economic model derived from a meta-analysis of randomized controlled trials in heart failure. European Journal of Heart Failure. 2011;13(4):450-9.

113. Laramée P, Wonderling D, Swain S, Al-Mohammad A, Mant J. Cost-effectiveness analysis of serial measurement of circulating natriuretic peptide concentration in chronic heart failure. Heart. 2013;99(4):267-71.

114. Miller G, Randolph S, Forkner E, Smith B, Galbreath AD. Long-term cost-effectiveness of disease management in systolic heart failure. Medical Decision Making. 2009.

115. Moertl D, Steiner S, Coyle D, Berger R. Cost-utility analysis of NT-proBNP-guided multidisciplinary care in chronic heart failure. International journal of technology assessment in health care. 2013;29(01):3-11.

116. Morimoto T, Hayashino Y, Shimbo T, Izumi T, Fukui T. Is B-type natriuretic peptide-guided heart failure management cost-effective? International journal of cardiology. 2004;96(2):177-81.

117. Pandor A, Thokala P, Gomersall T, Baalbaki H, Stevens J, Wang J, et al. Home telemonitoring or structured telephone support programmes after recent discharge in patients with heart failure: systematic review and economic evaluation. 2013.

118. National Clinical Guideline Centre. (2010). Chronic heart failure: the management of chronic heart failure in adults in primary and secondary care. London: National Clinical Guideline Centre. Available from: http://guidance.nice.org.uk/CG108/Guidance/pdf/English.

119. Galbreath AD, Krasuski RA, Smith B, Stajduhar KC, Kwan MD, Ellis R, et al. Long-term healthcare and cost outcomes of disease management in a large, randomized, community-based population with heart failure. Circulation. 2004;110(23):3518-26.

120. Paul SD, Kuntz KM, Eagle KA, Weinstein MC. Costs and effectiveness of angiotensin converting enzyme inhibition in patients with congestive heart failure. Archives of internal medicine. 1994;154(10):1143-9.

121. Delea TE, Vera-Llonch M, Richner RE, Fowler MB, Oster G. Cost effectiveness of carvedilol for heart failure. The American journal of cardiology. 1999;83(6):890-6.

122. Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. International journal of technology assessment in health care. 2008;24(03):244-58.

123. Goeree R, Burke N, O'Reilly D, Manca A, Blackhouse G, Tarride J-E. Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. Current Medical Research and Opinion®. 2007;23(4):671-82.

124. Angermann CE, Störk S, Gelbrich G, Faller H, Jahns R, Frantz S, et al. Mode of action and effects of standardized collaborative disease management on mortality and morbidity in patients with systolic heart failure: the Interdisciplinary Network for Heart Failure (INH) study. Circulation: Heart Failure. 2011:CIRCHEARTFAILURE. 111.962969.

125. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. New England Journal of Medicine. 2010;363(24):2301-9.

126. Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH, Investigators T-H. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. Journal of the American College of Cardiology. 2005;45(10):1654-64.

127. Drummond MF, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. Bmj. 1996;313(7052):275-83.

128. Eddy D. Technology assessment: the role of mathematical modeling. Assessing medical technologies. 1985:144-75.

129. Goehler A, Geisler BP, Manne JM, Jahn B, Conrads-Frank A, Gazelle GS, et al. Decision-Analytic Models to Simulate Health Outcomes and Costs in Heart Failure. Pharmacoeconomics. 2011;29(9):753-69.

130. Frederix GW, Afzali HHA, Dasbach EJ, Ward RL. Development and Use of Disease-Specific (Reference) Models for Economic Evaluations of Health Technologies: An Overview of Key Issues and Potential Solutions. PharmacoEconomics. 2015:1-5.

131. Handels RL, Wolfs CA, Aalten P, Joore MA, Verhey FR, Severens JL. Diagnosing Alzheimer's disease: A systematic review of economic evaluations. Alzheimer's & Dementia. 2014;10(2):225-37.

132. Getsios MD, Blume S, Ishak KJ, Maclaine GD. Cost effectiveness of donepezil in the treatment of mild to moderate Alzheimer's disease. Pharmacoeconomics. 2010;28(5):411-27.

133. Van Gestel A, Severens JL, Webers CA, Beckers HJ, Jansonius NM, Schouten JS. Modeling complex treatment strategies: construction and validation of a discrete event simulation model for glaucoma. Value in Health. 2010;13(4):358-67.

134. Rennie D, Luft HS. Pharmacoeconomic analyses: making them transparent, making them credible. JAMA. 2000;283(16):2158-60.

135. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. BMC medicine. 2013;11(1):80.

136. Raphael C, Briscoe C, Davies J, Whinnett ZI, Manisty C, Sutton R, et al. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart. 2007;93(4):476-82.

137. Lewis EF, Li Y, Pfeffer MA, Solomon SD, Weinfurt KP, Velazquez EJ, et al. Impact of cardiovascular events on change in quality of life and utilities in patients after myocardial infarction: a VALIANT study (valsartan in acute myocardial infarction). JACC: Heart Failure. 2014;2(2):159-65.

138. Yao G, Freemantle N, Flather M, Tharmanathan P, Coats A, Poole-Wilson PA. Long-Term Cost-Effectiveness Analysis of Nebivolol Compared with Standard Care in Elderly Patients with Heart Failure. Pharmacoeconomics. 2008;26(10):879-89.

139. Alla F, Briançon S, Guillemin F, Juillière Y, Mertès PM, Villemot JP, et al. Self-rating of quality of life provides additional prognostic information in heart failure. Insights into the EPICAL study. European Journal of Heart Failure. 2002;4(3):337-43.

140. Ramos S, Prata J, Rocha-Gonçalves F, Bettencourt P, Coelho R. Quality of Life Predicts Survival and Hospitalisation in a Heart Failure Portuguese Population. Applied Research in Quality of Life. 2017;12(1):35-48.

141. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Quality of Life Research. 1993;2(3):169-80.

142. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.

143. Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. Health and Quality of Life Outcomes. 2010;8(1):13.

144. Calvert MJ, Freemantle N, Cleland JGF. The impact of chronic heart failure on health-related quality of life data acquired in the baseline phase of the CARE-HF study. European Journal of Heart Failure. 2005;7(2):243-51.

145. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Nederlands tijdschrift voor geneeskunde. 2005;149(28):1574-8.

146. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

147. Rich MW, Beckham V, Wittenberg C, Leven CL, Freedland KE, Carney RM. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. New England Journal of Medicine. 1995;333(18):1190-5.

148. Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, et al. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. Archives of internal medicine. 1997;157(1):99-104.

149. Hasan A, Paul V. Telemonitoring in chronic heart failure. European heart journal. 2011;32(12):1457-64.

150. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoeconomics. 2018;36(6):675-97.

151. Böckerman P, Johansson E, Saarni SI. Do established health-related quality-of-life measures adequately capture the impact of chronic conditions on subjective well-being? Health Policy. 2011;100(1):91-5.

152. Bosworth H, Steinhauser K, Orr M, Lindquist J, Grambow S, Oddone E. Congestive heart failure patients' perceptions of quality of life: the integration of physical and psychosocial factors. Aging & mental health. 2004;8(1):83-91.

153. Heo S, Lennie TA, Okoli C, Moser DK. Quality of life in patients with heart failure: ask the patients. Heart & Lung: The Journal of Acute and Critical Care. 2009;38(2):100-8.

154. Blinderman CD, Homel P, Billings JA, Portenoy RK, Tennstedt SL. Symptom distress and quality of life in patients with advanced congestive heart failure. Journal of pain and symptom management. 2008;35(6):594-603.

155. Jaarsma T, Halfens R, Tan F, Abu-Saad HH, Dracup K, Diederiks J. Self-care and quality of life in patients with advanced heart failure: the effect of a supportive educational intervention. Heart & Lung: The Journal of Acute and Critical Care. 2000;29(5):319-30.

156. Harrison MB, Browne GB, Roberts J, Tugwell P, Gafni A, Graham ID. Quality of life of individuals with heart failure: a randomized trial of the effectiveness of two models of hospital-to-home transition. Medical care. 2002;40(4):271-82.

157. Caro JJ, Migliaccio-Walle K, O'Brien JA, Nova W, Kim J, Hauch O, et al. Economic implications of extended-release metoprolol succinate for heart failure in the MERIT-HF trial: a US perspective of the MERIT-HF trial. Journal of cardiac failure. 2005;11(9):647-56.

158. de Ridder D, Schreurs K. Developing interventions for chronically ill patients: is coping a helpful concept? Clinical psychology review. 2001;21(2):205-40.

159. Rector T, Cohn J. Minnesota living with heart failure questionnaire. Minnesota: University of Minnesota. 2004.

160. De Craen AJ, Kaptchuk TJ, Tijssen JG, Kleijnen J. Placebos and placebo effects in medicine: historical overview. Journal of the Royal Society of Medicine. 1999;92(10):511-5.

161. Johansson B, Brandberg Y, Hellbom M, Persson C, Petersson L, Berglund G, et al. Health-related quality of life and distress in cancer patients: results from a large randomised study. British journal of cancer. 2008;99(12):1975-83.

162. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics—2013 update. Circulation. 2013;127(1):e6-e245.

163. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, et al. Heart disease and stroke statistics—2008 update. Circulation. 2008;117(4):e25-e146.

164. Kilgore M, Patel HK, Kielhorn A, Maya JF, Sharma P. Economic burden of hospitalizations of Medicare beneficiaries with heart failure. Risk management and healthcare policy. 2017;10:63.

165. Mueller C, Christ M, Cowie M, Cullen L, Maisel AS, Masip J, et al. European Society of Cardiology-Acute Cardiovascular Care Association Position paper on acute heart failure: A call for interdisciplinary care. European Heart Journal: Acute Cardiovascular Care. 2017;6(1):81-6.

166. Weintraub NL, Collins SP, Pang PS, Levy PD, Anderson AS, Arslanian-Engoren C, et al. Acute heart failure syndromes: emergency department presentation, treatment, and disposition: current approaches and future aims: a scientific statement from the American Heart Association. Circulation. 2010;122(19):1975-96.

167. Clark RA, Inglis SC, McAlister FA, Cleland JG, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. Bmj. 2007;334(7600):942.

168. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. BMJ Publishing Group Ltd and British Cardiovascular Society; 2016.

169. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of internal medicine. 2006;144(3):201-9.

170. Zhang J, Goode KM, Cuddihy PE, Cleland JG. Predicting hospitalization due to worsening heart failure using daily weight measurement: analysis of the Trans-European Network-Home-Care Management System (TEN-HMS) study. European journal of heart failure. 2009;11(4):420-7.

171. Ledwidge MT, O'hanlon R, Lalor L, Travers B, Edwards N, Kelly D, et al. Can individualized weight monitoring using the HeartPhone algorithm improve sensitivity for clinical deterioration of heart failure? European journal of heart failure. 2013;15(4):447-55.

172. Anand IS, Tang WW, Greenberg BH, Chakravarthy N, Libbus I, Katra RP. Design and performance of a multisensor heart failure monitoring algorithm: results from the multisensor monitoring in congestive heart failure (MUSIC) study. Journal of cardiac failure. 2012;18(4):289-95.

173. Gyllensten IC, Bonomi AG, Goode KM, Reiter H, Habetha J, Amft O, et al. Early indication of decompensated heart failure in patients on home-telemonitoring: a comparison of prediction algorithms based on daily weight and noninvasive transthoracic bio-impedance. JMIR medical informatics. 2016;4(1).

174. Gyllensten IC, Crundall-Goode A, Aarts RM, Goode KM. Simulated case management of home telemonitoring to assess the impact of different alert algorithms on work-load and clinical decisions. BMC medical informatics and decision making. 2017;17(1):11.

175. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267-88.

176. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann; 2016.

177. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. New England Journal of Medicine. 1975;293(5):229-34.

178. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database of Systematic Reviews. 2015(10).

179. Balk AH, Davidse W, Dommelen P, Klaassen E, Caliskan K, Burgh P, et al. Tele-guidance of chronic heart failure patients enhances knowledge about the disease. A multi-centre, randomised controlled study. European journal of heart failure. 2008;10(11):1136-42.

180. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. European journal of heart failure. 2012;14(3):333-40.

181. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation. 2011:CIRCULATIONAHA. 111.018473.

182. Lyngå P, Persson H, Hägg-Martinell A, Hägglund E, Hagerman I, Langius-Eklöf A, et al. Weight monitoring in patients with severe heart failure (WISH). A randomized controlled trial. European journal of heart failure. 2012;14(4):438-44.

183. Soran OZ, Piña IL, Lamas GA, Kelsey SF, Selzer F, Pilotte J, et al. A randomized clinical trial of the clinical effects of enhanced heart failure monitoring using a computer-based telephonic monitoring system in older minorities and women. Journal of cardiac failure. 2008;14(9):711-7.

184. Vuorinen A-L, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. Journal of medical Internet research. 2014;16(12).

185. Woodend AK, Sherrard H, Fraser M, Stuewe L, Cheung T, Struthers C. Telehome monitoring in patients with cardiac disease who are at high risk of readmission. Heart & Lung: The Journal of Acute and Critical Care. 2008;37(1):36-45.

186. Mortara A, Pinna GD, Johnson P, Maestri R, Capomolla S, La Rovere MT, et al. Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure). European journal of heart failure. 2009;11(3):312-8.

187. AAPC Coder. HCPCS Code G0406 Follow-up inpatient consultation, limited, physicians typically spend 15 minutes communicating with the patient via telehealth [03/05/2018]. Available from: https://coder.aapc.com/hcpcs-codes/G0406.

188. Centers for Medicare and Medicaid Services (CMS). License for Use of Current Procedural Terminology, Fourth Edition ("CPT®") [03/05/2018]. Available from: https://www.cms.gov/apps/physician-fee-schedule/license-agreement.aspx.

189. Basset A, Nowak E, Castellant P, Gut-Gobert C, Le Gal G, L'her E. Development of a clinical prediction score for congestive heart failure diagnosis in the emergency care setting: the Brest score. The American journal of emergency medicine. 2016;34(12):2277-83.

190. Steinhart BD, Levy P, Vandenberghe H, Moe G, Yan AT, Cohen A, et al. A Randomized Control Trial Using a Validated Prediction Model for Diagnosing Acute Heart Failure in Undifferentiated Dyspneic Emergency Department Patients—Results of the GASP4Ar Study. Journal of cardiac failure. 2017;23(2):145-52.

191. Cubbon R, Woolston A, Adams B, Gale C, Gilthorpe M, Baxter P, et al. Prospective development and validation of a model to predict heart failure hospitalisation. Heart. 2014;100(12):923-9.

192. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2015;109(4):285-90.

193. Lassus J, Gayat E, Mueller C, Peacock WF, Spinar J, Harjola V-P, et al. Incremental value of biomarkers to clinical variables for mortality prediction in acutely decompensated heart failure: the Multinational Observational Cohort on Acute Heart Failure (MOCA) study. International journal of cardiology. 2013;168(3):2186-94.

194. Shah AM, Claggett B, Sweitzer NK, Shah SJ, Anand IS, O'meara E, et al. Cardiac structure and function and prognosis in heart failure with preserved ejection fraction: findings from the echocardiographic study of the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) Trial. Circulation: Heart Failure. 2014:CIRCHEARTFAILURE. 114.001583.

195. Shah KB, Kop WJ, Christenson RH, Diercks DB, Kuo D, Henderson S, et al. Post-discharge changes in NT-proBNP and quality of life after acute dyspnea hospitalization as predictors of one-year outcomes. Clinical biochemistry. 2010;43(18):1405-10.

196. Steinhart B, Thorpe KE, Bayoumi AM, Moe G, Januzzi JL, Mazer CD. Improving the diagnosis of acute heart failure using a validated prediction model. Journal of the American College of Cardiology. 2009;54(16):1515-21.

197. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Australian Journal of Physiotherapy. 2006;52(3):157-63.

198. Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? : Springer; 2003.

199. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. Health care management science. 1998;1(2):133-40.

200. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics. 2006;24(11):1043-53.

201. Heeg BM, Damen J, Buskens E, Caleo S, de Charro F, van Hout BA. Modelling approaches. Pharmacoeconomics. 2008;26(8):633-48.

202. Tran-Duy A, Boonen A, van de Laar MA, Franke AC, Severens JL. A discrete event modelling framework for simulation of long-term outcomes of sequential treatment strategies for ankylosing spondylitis. Annals of the rheumatic diseases. 2011;70(12):2111-8.

203. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. Health economics. 2003;12(10):837-48.

204. Eddy DM. Accuracy versus transparency in pharmacoeconomic modelling. Pharmacoeconomics. 2006;24(9):837-44.

205. Grustam AS, Severens JL, De Massari D, Buyukkaramikli N, Koymans R, Vrijhoef HJ. Cost-effectiveness analysis in telehealth: a comparison between home telemonitoring, nurse telephone support, and usual care in chronic heart failure management. Value in Health. 2018;21(7):772-82.

206. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield: Report by the Decision Support Unit. 2011.

207. Alehagen U, Rahmqvist M, Paulsson T, Levin LÅ. Quality-adjusted life year weights among elderly patients with heart failure. European journal of heart failure. 2008;10(10):1033-9.

208. Braunstein JB, Anderson GF, Gerstenblith G, Weller W, Niefeld M, Herbert R, et al. Noncardiac comorbidity increases preventable hospitalizations and mortality among Medicare beneficiaries with chronic heart failure. Journal of the American College of Cardiology. 2003;42(7):1226-33.

209. Statistics Netherlands. CBS Statline - Annual change in consumer price index; from 1963 2020 [13/08/2020]. Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?fromstatweb.

210. Zorginstituut Nederland. Guidelines for conducting economic evaluations in healthcare [in Dutch: Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg] 2016. Diemen: Zorginstituut Nederland. 2016.

211. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan SS. Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg In opdracht van Zorginstituut Nederland Geactualiseerde versie. 2015.

212. Stevanovic J, Denee L, Koenders J, Postma M. Incidence Description and Costs of Acute Heart Failure in the Netherlands. Value in Health. 2014;17(7):A328.

213. Hoogendoorn M, Ramos IC, Baldwin M, Guix NG-R, Rutten-van Mölken MP. Broadening the perspective of cost-effectiveness modeling in chronic obstructive pulmonary disease: a new patient-level simulation model suitable to evaluate stratified medicine. Value in Health. 2019;22(3):313-21.

214. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

215. Vemer P, Ramos IC, Van Voorn G, Al M, Feenstra T. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. 2016;34(4):349-61.

216. Seferovic PM, Ponikowski P, Anker SD, Bauersachs J, Chioncel O, Cleland JG, et al. Clinical practice update on heart failure 2019: pharmacotherapy, procedures, devices and patient management. An expert consensus meeting report of the Heart Failure Association of the European Society of Cardiology. European journal of heart failure. 2019;21(10):1169-86.

217. O'connor CM, Abraham WT, Albert NM, Clare R, Stough WG, Gheorghiade M, et al. Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). American heart journal. 2008;156(4):662-73.

218. Ford I, Robertson M, Komajda M, Böhm M, Borer JS, Tavazzi L, et al. Top ten risk factors for morbidity and mortality in patients with chronic systolic heart failure and elevated heart rate: the SHIFT Risk Model. International journal of cardiology. 2015;184:163-9.

219. Pocock SJ, Wang D, Pfeffer MA, Yusuf S, McMurray JJ, Swedberg KB, et al. Predictors of mortality and morbidity in patients with chronic heart failure. European heart journal. 2006;27(1):65-75.

220. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. Medical Decision Making. 2012;32(5):690-700.

221. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? Value in health. 2010;13(8):1056-60.

222. Willis M, Fridhammar A, Gundgaard J, Nilsson A, Johansen P. Comparing the Cohort and Micro-Simulation Modeling Approaches in Cost-Effectiveness Modeling of Type 2 Diabetes Mellitus: A Case Study of the IHE Diabetes Cohort Model and the Economics and Health Outcomes Model of T2DM. PharmacoEconomics. 2020:1-17.

223. Caro JJ. Pharmacoeconomic analyses using discrete event simulation. Pharmacoeconomics. 2005;23(4):323-32.

224. Caro JJ, Möller J. Advantages and disadvantages of discrete-event simulation for health economic analyses. Expert review of pharmacoeconomics & outcomes research. 2016;16(3):327-9.

225. Zhang J, Goode KM, Cuddihy PE, Cleland JG, Investigators TH. Predicting hospitalization due to worsening heart failure using daily weight measurement: analysis of the Trans-European Network-Home-Care Management System (TEN-HMS) study. European journal of heart failure. 2009;11(4):420-7.

226. Anand IS, Tang WW, Greenberg BH, Chakravarthy N, Libbus I, Katra RP, et al. Design and performance of a multisensor heart failure monitoring algorithm: results from the multisensor monitoring in congestive heart failure (MUSIC) study. Journal of cardiac failure. 2012;18(4):289-95.

227. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing.... presumed at random: cost-analysis of incomplete data. Health economics. 2003;12(5):377-92.

228. Gomes M, Díaz-Ordaz K, Grieve R, Kenward MG. Multiple imputation methods for handling missing data in cost-effectiveness analyses that use data from hierarchical studies: an application to cluster randomized trials. Medical decision making. 2013;33(8):1051-63.

229. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in analysing utility data for cost-utility analysis. Pharmacoeconomics. 2015;33(4):355-66.

230. Hughes D, Charles J, Dawoud D, Edwards RT, Holmes E, Jones C, et al. Conducting economic evaluations alongside randomised trials: current methodological issues and novel approaches. Pharmacoeconomics. 2016;34(5):447-61.

231. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32(12):1157-70.

232. Devin Incerti. Parametric survival modeling 2019. Available from: https://devinincerti.com/2019/06/18/parametric_survival.html.

233. Volksgezondheidenzorg. Hartfalen - Cijfers & Context - Huidige situatie 2020 [06/01/2021]. Available from: https://www.volksgezondheidenzorg.info/onderwerp/hartfalen/cijfers-context/huidige-situatie#node-nieuwe-gevallen-hartfalen-huisartsenpraktijk.

234. Maric B, Kaan A, Ignaszewski A, Lear SA. A systematic review of telemonitoring technologies in heart failure. European journal of heart failure. 2009;11(5):506-17.

235. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane database of systematic reviews. 2010(8).

236. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. Bmj. 2012;344:e3874.

237. Henderson C, Knapp M, Fernández J-L, Beecham J, Hirani SP, Cartwright M, et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth

questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. Bmj. 2013;346:f1035.

238. Albuquerque de Almeida F, Corro Ramos I, Rutten-van Mölken M, Al M. Modeling Early Warning Systems: Construction and Validation of a Discrete Event Simulation Model for Heart Failure. Value in Health.

239. Zorginstituut Nederland. Ziektelast in de praktijk - De theorie en praktijk van het berekenen van ziektelast bij pakketbeoordelingen 2018 [06/01/2021]. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2018/05/07/ziektelast-in-de-praktijk/Ziektelast+in+de+praktijk_definitief.pdf.

240. Stolk EA, van Donselaar G, Brouwer WB, Busschbach JJ. Reconciliation of economic concerns and health policy. Pharmacoeconomics. 2004;22(17):1097-107.

241. Versteegh MM, Ramos IC, Buyukkaramikli NC, Ansaripour A, Reckers-Droog VT, Brouwer WB. Severity-Adjusted Probability of Being Cost Effective. PharmacoEconomics. 2019;37(9):1155-63.

242. Institute for Medical Technology Assessment: iMTA. The iMTA Disease Burden Calculator (iDBC) for absolute and shortfall [06/01/2021]. Available from: https://imta.shinyapps.io/iDBC/.

243. Centraal Bureau voor de Statistiek (CBS). Landelijke Basisregistratie Ziekenhuiszorg 2020 [06/01/2021]. Available from: https://bronnen.zorggegevens.nl/Bron?naam=Landelijke-Basisregistratie-Ziekenhuiszorg.

244. Koulaouzidis G, Iakovidis D, Clark A. Telemonitoring predicts in advance heart failure admissions. International journal of cardiology. 2016;216:78-84.

245. Z-index. 2019 [06/01/2021]. Available from: https://www.z-index.nl.

246. Kanters TA, Bouwmans CA, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. PLoS One. 2017;12(11):e0187477.

247. Ramos IC, Versteegh MM, de Boer RA, Koenders JM, Linssen GC, Meeder JG, et al. Cost effectiveness of the angiotensin receptor neprilysin inhibitor sacubitril/valsartan for patients with chronic heart failure and reduced ejection fraction in the Netherlands: a country adaptation analysis under the former and current Dutch Pharmacoeconomic Guidelines. Value in Health. 2017;20(10):1260-9.

248. Evans M, Moes RG, Pedersen KS, Gundgaard J, Pieber TR. Cost-Effectiveness of Insulin Degludec Versus Insulin Glargine U300 in the Netherlands: Evidence From a Randomised Controlled Trial. Advances in therapy. 2020;37(5):2413-26.

249. van Baal PH, Wong A, Slobbe LC, Polder JJ, Brouwer WB, de Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. Pharmacoeconomics. 2011;29(3):175-87.

250. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health. 2010;13(5):509-18.

251. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health economics. 1994;3(2):95-104.

252. Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2004;66(3):751-69.

253. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value in Health. 2005;8(1).

254. Corro Ramos I, Hoogendoorn M, Rutten-van Mölken MP. How to Address Uncertainty in Health Economic Discrete-Event Simulation Models: An Illustration for Chronic Obstructive Pulmonary Disease. Medical Decision Making. 2020;40(5):619-32.

255. Briggs AH. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics. 2000;17(5):479-500.

256. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. Medical decision making. 1985;5(2):157-77.

257. Al MJ. Cost-effectiveness acceptability curves revisited. Pharmacoeconomics. 2013;31(2):93-100.

258. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of health economics. 1999;18(3):341-64.

259. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). The Lancet. 2002;360(9334):711-5.

260. Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Volksgezondheid Toekomst Verkenning. 2018.

261. Linssen G, Veenis JF, Brunner-La Rocca HP, van Pol P, Engelen D, van Tooren R, et al. Differences in guideline-recommended heart failure medication between Dutch heart failure clinics: an analysis of the CHECK-HF registry. Netherlands Heart Journal. 2020;28:334-44.

262. Boyne JJ, Di Van Asselt A, Gorgels AP, Steuten LM, De Weerd G, Kragten J, et al. Cost–effectiveness analysis of telemonitoring versus usual care in patients with heart failure: The TEHAF–study. Journal of telemedicine and telecare. 2013;19(5):242-8.

263. Boyne JJ, Vrijhoef HJ, De Wit R, Gorgels AP. Telemonitoring in patients with heart failure, the TEHAF study: Study protocol of an ongoing prospective randomised trial. International journal of nursing studies. 2011;48(1):94-9.

264. Boyne JJ, Vrijhoef HJ, Crijns HJ, De Weerd G, Kragten J, Gorgels AP, et al. Tailored telemonitoring in patients with heart failure: results of a multicentre randomized controlled trial. European journal of heart failure. 2012;14(7):791-801.

265. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database of Systematic Reviews. 2015(10).

266. European Medicines Agency (EMA). Medical Devices [30/12/2020]. Available from: https://www.ema.europa.eu/en/human-regulatory/overview/medical-devices.

267. European Patent Office (EPO). Annual reports [30/12/2012]. Available from: https://www.epo.org/about-us/annual-reports-statistics/annual-report.html.

268. MedTech Europe. The European Medical Technology Industry in figures 2020. 2020.

269. Fargen KM, Frei D, Fiorella D, McDougall CG, Myers PM, Hirsch JA, et al. The FDA approval process for medical devices: an inherently flawed system or a valuable pathway for innovation? Journal of neurointerventional surgery. 2013;5(4):269-75.

270. Sorenson C, Drummond M. Improving medical device regulation: the United States and Europe in perspective. The Milbank Quarterly. 2014;92(1):114-50.

271. Kramer DB, Xu S, Kesselheim AS. How does medical device regulation perform in the United States and the European union? A systematic review. PLoS Med. 2012;9(7):e1001276.

272. Fattore G, Maniadakis N, Mantovani LG, Boriani G. Health technology assessment: what is it? Current status and perspectives in the field of electrophysiology. Europace. 2011;13(suppl_2):ii49-ii53.

273. Schreyögg J, Bäumler M, Busse R. Balancing adoption and affordability of medical devices in Europe. Health Policy. 2009;92(2-3):218-24.

274. Taylor RS, Iglesias CP. Assessing the clinical and cost-effectiveness of medical devices and drugs: are they that different? Value in Health. 2009;12(4):404-6.

275. Eikermann M, Gluud C, Perleth M, Wild C, Sauerland S, Gutierrez-Ibarluzea I, et al. Commentary: Europe needs a central, transparent, and evidence based regulation process for devices. Bmj. 2013;346.

276. European Union. Types of EU law [02/01/2021]. Available from: https://ec.europa.eu/info/law/law-making-process/types-eu-law_en#primary-vs-secondary-law.

277. Craig P, De Búrca G. EU law: text, cases, and materials: Oxford university press; 2020.

278. Bradford A. The Brussels effect: How the European Union rules the world: Oxford University Press, USA; 2020.

279. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use (OJ L 311, 28.11.2001, p. 67–128), (Medicinal Products Directive).

280. Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (Text with EEA relevance) (OJ L 136, 30.4.2004, p. 1–33), (Medicinal Products Regulation).

281. U.S. National Library of Medicine. ClinicalTrials.gov [23/12/2020]. Available from: https://clinicaltrials.gov/.

282. Thokala P, Carlson JJ, Drummond M. HTA'd in the USA: A Comparison of ICER in the United States with NICE in England and Wales. Journal of Managed Care & Specialty Pharmacy. 2020;26(9):1162-70.

283. Haute Autorité de Santé (HAS). [23/12/2020]. Available from: https://www.has-sante.fr/.

284. Gemeinsamer Bundesausschuss (G-BA). [23/12/2020]. Available from: https://www.g-ba.de/.

285. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). [23/12/2020]. Available from: https://www.iqwig.de/.

286. Agenzia Italiana del Farmaco (AIFA). [23/12/2020]. Available from: https://www.aifa.gov.it/en/web/guest/home.

287. Agenzia Nazionale per i Servizi Sanitari Regionali (AGENAS). [23/12/2020]. Available from: https://www.agenas.gov.it/.

288. National Institute for Health and Care Excellence (NICE). [23/12/2020]. Available from: https://www.nice.org.uk/.

289. Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS). [23/12/2020]. Available from: https://redets.sanidad.gob.es/.

290. Agencia Española de Medicamentos y Productos Sanitarios (AEMPS). [23/12/2020]. Available from: https://www.aemps.gob.es/.

291. COUNCIL DIRECTIVE of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices (90/385/EEC) (OJ L 189, 20.7.1990, p. 17), (Active Implantable Medical Devices Directive).

292. COUNCIL DIRECTIVE 93/42/EEC of 14 June 1993 concerning medical devices (OJ L 169, 12.7.1993, p. 1), (Medical Devices Directive).

293. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 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 (Text with EEA relevance.), (Medical Devices Regulation).

294. Regulation (EU) 2020/561 of the European Parliament and of the Council of 23 April 2020 amending Regulation (EU) 2017/745 on medical devices, as regards the dates of application of certain of its provisions (Text with EEA relevance), PE/10/2020/REV/1, (OJ L 130, 24.4.2020, p. 18–22), (Medical Devices Directive Amendment).

295. French-Mowat E, Burnett J. How are medical devices regulated in the European Union? Journal of the Royal Society of Medicine. 2012;105(1_suppl):22-8.

296. REGULATION (EU) No 536/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, (Clinical Trials Regulation).

297. de Morpurgo M. The E uropean U nion as a Global Producer of Transnational Law of Risk Regulation: A Case Study on Chemical Regulation. European Law Journal. 2013;19(6):779-98.

298. Naci H, Salcher-Konrad M, Kesselheim AS, Wieseler B, Rochaix L, Redberg RF, et al. Generating comparative evidence on new drugs and devices before approval. The Lancet. 2020;395(10228):986-97.

299. Fraser AG, Butchart EG, Szymański P, Caiani EG, Crosby S, Kearney P, et al. The need for transparency of clinical evidence for medical devices in Europe. The Lancet. 2018;392(10146):521-30.

300. Dhruva SS, Bero LA, Redberg RF. Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices. Jama. 2009;302(24):2679-85.

301. Rathi VK, Krumholz HM, Masoudi FA, Ross JS. Characteristics of clinical studies conducted over the total product life cycle of high-risk therapeutic medical devices receiving FDA premarket approval in 2010 and 2011. Jama. 2015;314(6):604-12.

302. Chen CE, Dhruva SS, Redberg RF. Inclusion of comparative effectiveness data in high-risk cardiovascular device studies at the time of premarket approval. JAMA. 2012;308(17):1740-2.

303. Commission welcomes the move towards more innovative health technologies for patients (IP/21/3142) [press release]. 2021.

304. Garrett Z. An Analysis Of HTA And Reimbursement Processes In Eunethta Partner Countries: Implications For Implementing HTA Cooperation. Value in Health. 2017;20(9):A678.

305. Zhao Y, Feng H-m, Qu J, Luo X, Ma W-j, Tian J-h. A systematic review of pharmacoeconomic guidelines. Journal of medical economics. 2018;21(1):85-96.

306. Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs. Same or different? 2009.

307. Tarricone R, Torbica A, Drummond M, Group MP. Key recommendations from the MedtecHTA project. Health economics. 2017;26:145-52.

308. Schnell-Inderst P, Hunger T, Conrads-Frank A, Arvandi M, Siebert U. Ten recommendations for assessing the comparative effectiveness of therapeutic medical devices: a targeted review and adaptation. Journal of clinical epidemiology. 2018;94:97-113.

309. EUnetHTA JA2. Guideline "Therapeutic medical devices". 2015.

310. Henschke C, Panteli D, Perleth M, Busse R. Taxonomy of medical devices in the logic of health technology assessment. 2015.

311. United Nations (UN). Early Warning Systems [18/04/2021]. Available from: https://www.un.org/en/climatechange/climate-solutions/early-warning-systems.

312. Bussiere M, Fratzscher M. Towards a new early warning system of financial crises. journal of International Money and Finance. 2006;25(6):953-73.

313. United Nations Development Programme (UNDP). Five approaches to build functional early warning systems. 2018.

314. Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. Journal of the American College of Cardiology. 2005;45(10):1654-64.

315. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

316. Perera WR, Hurst JR, Wilkinson TM, Sapsford RJ, Müllerova H, Donaldson GC, et al. Inflammatory changes, recovery and recurrence at COPD exacerbation. European Respiratory Journal. 2007;29(3):527-34.

317. Toy EL, Gallagher KF, Stanley EL, Swensen AR, Duh MS. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2010;7(3):214-28.

318. Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: stepping forward towards optimal decision rules. Health economics. 2007;16(3):307-17.

319. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, Van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. Health economics. 2019;28(1):87-100.

320. Keeney RL, Raiffa H, Meyer RF. Decisions with multiple objectives: preferences and value trade-offs: Cambridge university press; 1993.

321. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making—an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value in health. 2016;19(1):1-13.

322. Baltussen R, Marsh K, Thokala P, Diaby V, Castro H, Cleemput I, et al. Multicriteria decision analysis to support health technology assessment agencies: benefits, limitations, and the way forward. Value in Health. 2019;22(11):1283-8.

323. Bundesinstitut für Arzneimittel und Medizinprodukte. Digital health applications (DiGA) [08/06/2021]. Available from: https://www.bfarm.de/DE/Medizinprodukte/DVG/_node.html.