Improving Telemonitoring for Chronic Heart Failure Patients

A Data-driven Investigation into improving Equity and AI models for Telemedicine

Xylander, Alexander Arndt Pasgaard

DOI (link to publication from Publisher):
10.54337/aau682324294

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

IMPROVING TELEMONITORING FOR CHRONIC HEART FAILURE

A DATA-DRIVEN INVESTIGATION INTO IMPROVING EQUITY AND AI MODELS FOR TELEMEDICINE INTERVENTIONS

BY
ALEXANDER ARNDT PASGAARD XYLANDER

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

IMPROVING TELEMONITORING FOR CHRONIC HEART FAILURE

A DATA-DRIVEN INVESTIGATION INTO IMPROVING EQUITY AND AI MODELS FOR TELEMEDICINE INTERVENTIONS

by

Alexander Arndt Pasgaard Xylander

AALBORG UNIVERSITY
DENMARK

Dissertation submitted

Saturday, December 2, 2023

Dissertation submitted:	December 2, 2023
PhD supervisor:	Associate Professor Flemming Witt Udsen Aalborg University
Assistant PhD supervisors:	Professor Ole Hejlesen Aalborg University Associate Professor Simon Lebech Cichosz Aalborg University Associate Professor Morten Hasselstrøm Jensen Aalborg University
PhD committee:	Associate Professor Lasse Riis Østergaard (chair) Aalborg University, Denmark Professor Abdollah Vahabpour-Roudsari University of Victoria, Canada Professor Mariann Fossum University of Agder, Norway
PhD Series:	Faculty of Medicine, Aalborg University
Department:	Department of Health Science and Technology

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-588-4

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2024

CV

Alexander Arndt Pasgaard Xylander has a master's degree in public health, cand.scient.san.publ, from Aalborg University obtained in 2018. Prior to his PhD enrollment, Alexander worked on several research projects focused on advanced analysis of health care data. During his master's degree, his research focused on an epidemiological analysis of inequity in health care utilization. After finishing his degree, Alexander was employed as a research assistant at the Danish Center for Health Care Improvements, where he engaged in research tasks related to data collection, registry data applications and teaching duties related to health economics research. During his tenure as a research assistant, Alexander further developed his methodological skills with a particular emphasis on machine learning and artificial intelligence-related modeling. In 2020, Alexander was enrolled as a PhD Student in the "Biomedical Engineering and Neuroscience" (formerly "Clinical Science and Biomedicine") program at the Doctoral School in Medicine, Biomedical Science and Technology.

Alexander's PhD focused on the analysis and statistical modeling of deterioration for telemonitored chronic heart failure patients. Alexander's research was supervised by main supervisor Assistant Professor Flemming Witt Udsen along with assistant supervisors Professor Ole Hejlesen, Professor Morten Hasselstrøm Jensen, and Assistant Professor Simon Lebech Cichosz.

During his PhD, Alexander collaborated with the Region of North Denmark in developing new knowledge and technologies for the implementation of the Telecare Nord Heart Failure Telemedicine Project.

Alongside research duties, Alexander has fulfilled teaching duties at the Institute of Health, supervising public health, medicine, and biomedical engineering student projects as well as classroom teaching.

ENGLISH SUMMARY

This PhD thesis investigates the potential for improving telemonitoring of chronic heart failure (CHF) patients. This thesis is part of a project investigating supplementary telemedicine for chronic obstructive lung disease and CHF patients in the Region of Northern Denmark, “Telecare Nord”/ “Telecare Nord Heart Failure.”

The first part of the thesis presents the etiology and epidemiology of CHF and elaborates on the current trend of using telehealth technologies in the management of CHF patients. The first study, “Does geographical distance impact the effectiveness of supplementary telemedicine for chronic heart failure patients”, supports this section by investigating the literature on disparities in health care access for CHF patients and provides quantitative evidence that supplementary telemedicine can help address some of these inequities by mitigating geographical distance as a barrier to healthcare access.

The next part of the PhD investigates the potential for improving the telemonitoring strategies of CHF patients receiving telemedicine in the Region of Northern Denmark.

The second study, “Weekly Blood Pressure and Weight Measurements shows Potential for Predicting Deterioration in Telemonitored Chronic Heart Failure Patients”, is an exploratory study that aims to investigate whether weekly measurements of biometrical data, including blood pressure, weight, and pulse, can be used to discriminate between stable periods and periods leading to hospitalization for chronic heart failure patients through a paired observation strategy with 33 event and nonevent periods. The study finds that features based on changes in periodical trends of biometric values increase the discriminatory performance of the model.

The third study, “Prediction of 14-day hospitalization risk in chronic heart failure patients, using interpretable machine learning methods”, builds on the findings of the second study and uses machine learning algorithms to develop predictive models that provide risk estimates for nonelective hospitalizations in a 14-day time period. The study uses a supervised learning framework and includes 11,575 observations of

telemonitored values with a label indicating the presence of a nonelective hospitalization in the 14 days following the observation. The machine learning techniques included in the modeling have been chosen for their interpretability to improve eventual adaptation from the carers responsible for the monitoring of CHF patients.

While the overall performance of the developed predictive models is modest, the studies demonstrate that the more interpretable logistic regression model outperforms more complex machine learning algorithms on this dataset, suggesting that future research might be able to avoid black box models and the associated skepticisms and nonadaptation by clinicians.

DANSK RESUME

Denne artikel baserede ph.d.-afhandling undersøger potentialet for at forbedre telemonitorering af patienter med kronisk hjertesvigt / hjerteinsufficiens (CHF). Første del af afhandlingen præsenterer årsagerne og epidemiologien for CHF og uddyber den nuværende tendens til at anvende telemedicinteknologier i håndteringen af CHF-patienter. Den første artikel, "Does geographical distance impact the effectiveness of supplementary telemedicine for chronic heart failure patients", støtter denne sektion ved at undersøge litteraturen om uligheder i adgangen til sundhedspleje for CHF-patienter og giver kvantitative beviser for, at supplementær telemedicin kan hjælpe med at tackle nogle af disse uligheder ved at reducere geografisk afstand som en hindring for sundhedsplejeadgang.

Den næste del af ph.d.-afhandlingen undersøger potentialet for at forbedre telemonitoreringsstrategierne for CHF-patienter, der modtager telemedicin i Region Nordjylland, i forbindelse med "Telecare Nord Hjertesvigt" projektet.

Anden artikel, " Weekly Blood Pressure and Weight Measurements shows Potential for Predicting Deterioration in Telemonitored Chronic Heart Failure Patients", er en eksplorativ undersøgelse af hvorvidt ugentlige målinger af biometriske data, herunder blodtryk, vægt og puls, kan bruges til at skelne mellem stabile perioder og perioder, der fører til indlæggelse. Dette gøres ved hjælp af en parret observationsstrategi med 33 begivenheds- og ikke-begivenhedsperioder. Undersøgelsen viser, at inklusion af ændringer i periodiske tendenser af biometriske værdier i algoritmen øger modellens diskriminerende præstation.

Den tredje artikel, " Prediction of 14-day hospitalization risk in chronic heart failure patients, using interpretable machine learning methods", bygger på resultaterne fra anden artikel og bruger maskinlæringsalgoritmer til at udvikle kliniske forudsigelsesmodeller, der giver et risikoestimat for ikke-planlagte indlæggelser inden for den næste 14-dages periode for hver hjemmemåling udført af patienten. I artiklen anvendes et "supervised learning" design og den inkluderer 11.575 observationer af teleovervågede værdier med et label, der angiver tilstedeværelsen af en ikke-planlagt indlæggelse i de følgende 14 dage efter observationen. Maskinlæringsteknikkerne, der er inkluderet i modelleringen, er valgt for deres fortolkelighed for at understøtte klinisk implementering.

Afhandlingen er en del af et projekt, der undersøger supplementær telemedicin for patienter med kronisk obstruktiv lungesygdom og CHF-patienter i Region

Nordjylland, "Telecare Nord" / "Telecare Nord Hjertesvigt". Mens de udviklede forudsigelsesmodeller er utilstrækkeligt præcise til klinisk brug, så viser undersøgelserne, at den mere fortolkningsbare logistiske regressionsmodel overgår mere komplekse maskinlæringsalgoritmer på dette datasæt, hvilket antyder, at fremtidig forskning måske kan undgå sort-boks-modeller, og de dertilhørende udfordringer.

ACKNOWLEDGMENTS

I would like to thank a great many people for their help in making the completion of this thesis possible. A lot of both professional and personal challenges have occurred during my work on this project, from the very first day where a pandemic lockdown was instated to a divorce and subsequent changes in my personal life. I can honestly say that without the people around me and their support, I wouldn't've succeeded in conquering the challenges and difficulties inherent in such an endeavor.

I would like to thank all my supervisors, Flemming Witt Udsen, Ole Hejlesen, Simon Lebech Cichosz and Morten Hasselstrøm Jensen. For more than three years, they have supported and contributed to my work on the papers in this thesis. Whenever doubts about my own abilities or the validity of the project threatened to overwhelm my thoughts, my supervisors have offered their time, understanding and support in addition to their substantial expertise. I could ask for nothing more, and I will always be gracious for the kindness I have received.

I would also like to thank the rest of the Medical Informatics Group, who contributed to keeping my sanity mostly intact.

I would like to thank the Region of Northern Denmark and Ria Thaarup Høegh, who served as the primary collaborators in this project. Ria's enthusiasm for our research often surpasses our own, and while some interinstitutional projects get caught in bureaucratic quagmire, Ria has facilitated all of our requests with expediency and professionalism.

I would also like to thank the patients who contributed to the data used for this thesis. Their courage in engaging with a new healthcare technology is inspiring.

Last, I want to thank my family and friends. While the people who have supported me are too many to mention, my daughter, my brother and my girlfriend deserve special mention.

To Ellinor, without your steady supply of hugs and kisses, this thesis would not exist.

To Oliver, your dedication to conquering the challenges put before you inspire me, I am so proud to be your brother.

To Madeleine, the discovery of you surpasses that of any of my research.

TABLE OF CONTENTS

TABLE OF FIGURES

PREFACE

This text represents the PhD thesis submitted to the Doctoral School of Medicine at Aalborg University. The thesis is based upon three papers produced as a part of the PhD work conducted by Alexander Arndt Pasgaard Xylander in the years 2020-2023 at the Department of Health Science and Technology, Aalborg University.

The thesis is structured with both essay and paper sections. The first section is the background, which covers current knowledge and contemporary research in the field. This is followed by the overall objective of the thesis, as well as specific objectives for each of the included papers. The next section is a brief summary of each of the papers, for readers that might lack access to the papers themselves. Then, follow the papers (Papers II and III are replaced with placeholders for the published version due to copyright). The last sections of the thesis are a discussion of methodological considerations and impact of the findings and a conclusion.

The research and PhD work was conducted under the supervision of associate professor Flemming Witt Udsen and professor Ole Hejlesen, with additional support from associate professor Simon Lebech Cichosz and professor Morten Hasselstrøm Jensen.

LIST OF PUBLICATIONS

Thesis Papers

Paper I **Xylander AAP**, Cichosz SL, Hejlesen O, et al. Does geographical distance impact the effectiveness of supplementary telemedicine for Chronic Heart Failure patients.

Submitted and currently under review at BMC Public Health, 2023

Paper II **Xylander AAP**, Udsen FW, Jensen MH, et al. Weekly Blood Pressure and Weight Measurements shows Potential for Predicting Deterioration in Telemonitored Chronic Heart Failure Patients.

Submitted and currently under review at Journal of Telemedicine and Telecare, 2023

Paper III **Xylander AAP**, Cichosz SL, Jensen MH, et al. Prediction of 14-day hospitalization risk in chronic heart failure patients, using interpretable machine learning methods.

Submitted and currently under review at Health and Technology, 2023

Other papers by the author

Mæhlisen MH, **Pasgaard AA**, Mortensen RN, et al. Perceived stress as a risk factor of unemployment: A register-based cohort study. *BMC Public Health* 2018; 18: 1–11.

Cichosz SL, **Xylander AAP**. A Conditional Generative Adversarial Network for Synthesis of Continuous Glucose Monitoring Signals. *J Diabetes Sci Technol* 2022; 16: 1220–1223.

Pasgaard AA, Mæhlisen MH, Overgaard C, et al. Social capital and frequent attenders in general practice: A register-based cohort study. *BMC Public Health* 2018; 18: 1–1

ABBREVIATIONS

ACE-I: Angiotensin-converting enzyme inhibitors
AI: Artificial intelligence
AUC: Area under the curve
ARNI: Angiotensin receptor-neprilysin inhibitors
CHF: Chronic heart failure
CI: Confidence interval (95%)
COPD: Chronic obstructive pulmonary disease
EQ5D: EuroQoL-5 Dimensions
HF: Heart failure
HFrEF: Heart failure with reduced ejection fraction
HFmrEF: Heart failure with mildly reduced ejection fraction
HFpEF: Heart failure with preserved ejection fraction
LVEF: Left ventricular ejection fraction
MACD: Moving average convergence divergence
MEML: Mixed effects machine learning
ML: Machine Learning
MRA: Mineralocorticoid receptor antagonists
NYHA: New York Heart Association
RAAS: Renin-angiotensin-aldosterone system
ROC: Receiver-operator-curve
RoT: Rule-of-thumb
RR: Relative Risk
SF-36: 36-item short form health survey
SGLT-2: Sodium-glucose cotransporter-2
TCNH: Telecare Nord Heart Failure

CHAPTER 1. INTRODUCTION

Chronic heart failure (CHF) is a progressive chronic disease that constitutes a significant public health burden. It is estimated that 1-2% of the adult population (1) has CHF, and the prevalence is expected to increase in the future (1–5). CHF is correlated with increased mortality and morbidity and significantly reduced quality of life for patients with the diagnosis (1,6–8). The combination of high prevalence and incidence rates with a high burden of disease, makes CHF a substantial burden on healthcare systems worldwide (9). The progressive nature of the disease means that continuous adjustments in the care and medications by the patient are needed for optimal care. In many countries, CHF patients are routinely monitored by the appropriate hospital department or primary care practitioner to prevent further deterioration of the underlying disease as well as acute decompensation events. Despite preventive care, CHF patients will often continue to worsen, leading to an increased incidence of acute hospitalizations. To address the need for close monitoring of patients as well as increasing costs in the health care sector, telemedicine has been proposed as a tool for supporting patient self-care and providing timely and accurate data to the responsible clinicians (10–12).

However, several aspects of the technology are still unclear:

Which patients benefit from receiving telemedicine?

Socioeconomically disadvantaged CHF patients have worse outcomes and prognoses than other patients (13–16) even in countries with universal healthcare (13). Many potential explanations exist for this relationship (17), but one important factor might be geographical barriers to healthcare access (18). Telemedicine has been assumed to largely solve this issue, but some authors are still skeptical (18).

How do clinicians and patients best utilize the new monitoring technologies?

CHF entails significant use of health care resources, which is a burden for individual patients and constitutes a significant expenditure for the health care system (1,19,20). Effective telemonitoring technologies might help prevent deterioration and hospitalization events (10,19,21,22) but rely on proper implementation as well as effective support to identify patients who need clinical intervention. One such tool is clinical decision support algorithms, which use patient data for risk assessment on an individual patient level, providing additional context for clinicians (21).

This thesis will try to (partly) answer these questions. The following Chapter 2 - Background elaborates on the points presented in the introduction and provides the necessary context for the later paper-based chapters, particularly the methodology surrounding the implementation of predictive algorithms. In Chapter 3 – Thesis objective, the objectives of this thesis are presented along with a brief presentation of the specific aims of the individual papers. In Chapter 4 – Summary of papers, an abstract of each of the thesis-defining papers is presented, with the full papers available in Chapter 5 (not available in the online version). In Chapter 6 – Discussion, the main findings of the papers are discussed and contextualized to the methodological choices of the papers and each other, which leads to the overall conclusion in Chapter 7 – Conclusion.

CHAPTER 2. BACKGROUND

In this chapter, I will introduce the necessary background knowledge to understand the thesis objectives and their raison d'être. The first subsection provides a very brief introduction to the function and anatomy of the heart. To provide just a tiny bit of context for the later chapters on CHF, much more comprehensive writings on the anatomy and physiology of the heart are available elsewhere (23).

2.1. THE HEART

“The heart of animals is the foundation of their life, the sovereign of everything within them, the sun of their microcosm, that upon which all growth depends, from which all power proceeds.” — William Harvey, An Anatomical Disquisition on the Motion of the Heart and Blood in Animals, 1628.

William Harvey was among the first physicians to deduce the true nature of the heart, that of a 2-chambered mechanical pump that powers and regulates the movement and pressure of both venous and arterial blood through the body.

Since his seminal work was published in 1628, researchers, physicians and others have made tremendous contributions toward our understanding of the physiological structures and functions of the heart, but the leading quote still holds true – the heart is essential for human life (24).

In the following section, a brief introduction to the anatomy and physiology of the heart will be presented to better explain the complex pathologies associated with CHF.

Figure 1 illustrates the basic anatomy of the heart. The heart consists of two chambers, the left and right ventricles. The left ventricle is responsible for pumping oxygen-saturated blood through the systemic circuit, while the right ventricle pumps oxygen-deprived blood through the pulmonary circuit. The ventricles operate as a two-pump-series system with a filling and a contraction phase. During the filling phase, blood enters the ventricular chambers through the atria, and during the contraction phase, the heart contracts and pumps the blood through the pulmonary and systemic circuits.

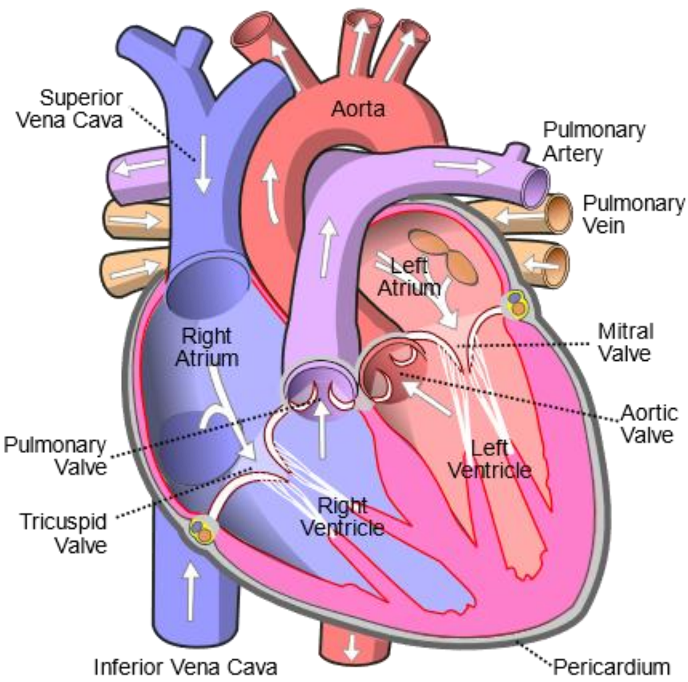


Figure 1 - Anatomy of the heart original illustration available at https://en.wikipedia.org/wiki/Atrium_%28heart%29

In healthy individuals, contraction of the heart is regulated by complex interactions between neurohormonal, hormonal and neuroelectric factors. This regulation is essential in maintaining a steady flow of oxygenated blood that is sufficient to address the needs of the body. When the oxygen requirements of the body increase, for example, during exercise, the heartbeat and respiratory frequency increase to compensate for the increased demand. In addition, the renin-angiotensin-aldosterone pathway is activated, which results in fluid retention and vasoconstriction; in healthy individuals, this improves the biomechanical properties of the cardiovascular system, increasing the capacity of the individual to maintain proper blood pressure. However, in individuals with heart failure, some of these mechanisms might result in a vicious cycle that results in paradoxical decreases in cardiac function, particularly in the long term. (23)

2.2. DEFINITION, ETIOLOGY AND CLASSIFICATION OF CHRONIC HEART FAILURE

When the heart is unable to fulfill the cardiovascular requirements of the body, a range of symptoms and pathological changes occur. This syndrome is referred to as heart failure (HF), and if the condition is irreversible, it is often referred to as chronic or congestive heart failure (CHF). In this thesis, HF and CHF are used interchangeably, with both referring to the chronic version of the disease, unless otherwise specified.

Due to the complex nature of the syndrome, several different definitions of HF exist in the medical research literature and guidelines (25). Some definitions, such as the recently proposed "Universal Definition of Heart Failure" (25), emphasize precise objective clinical criteria of decreased cardiac function, such as cardiac abnormalities, elevated natriuretic peptides and more; others, e.g., "the Framingham Criteria" (26), use a more conceptual definition of the syndrome, and some focus on the cardinal symptoms experienced by the patient (27). A recent collaboration between cardiological societies has proposed the following definition:

"A clinical syndrome with symptoms and/or signs caused by a structural and/or functional cardiac abnormality and corroborated by elevated natriuretic peptide levels and/or objective evidence of pulmonary or systemic congestion" (25)

While the purpose of this definition is to provide a standardized framework for working with HF, this definition still shows the challenging heterogeneity of HF syndrome between patients. Some patients experience symptoms of cardiovascular congestion, despite no or few objective clinical markers of disease, while others experience few symptoms despite elevated biomarkers and/or cardiac abnormalities. The symptoms and signs of CHF include (but is not limited to) (28):

Breathlessness
Oedema including ankle swelling
Fatigue
Chest pain
Coughing / wheezing
Tachycardia
Increased jugular venous pressure

With a wide range of conditions and diseases being potential causes of HF, the etiology is complex, with many different cardiovascular pathologies contributing to the development of the syndrome. Ischemic heart disease and hypertension are the most common causes of heart failure, particularly in Western and developed countries (5,29). Other potential causes include cardiac valve disease, cardiomyopathies, congenital heart disease, infective diseases, cancers and many more.

To provide a disease management strategy that is suited for the individual patient, HF classification schemas have been developed. Like the definition of the syndrome, multiple different classifications exist in the literature.

One classification schema commonly used in both research and clinical practice is the New York Heart Association functional classification system (NYHA). NYHA classifies HF patients into four different classes, I-IV, based on the perceived physical limitations and symptoms experienced by the patient.

1. HF patients with no limitations have a NYHA class of I.
2. HF patients who experience a slight limitation with undue shortness of breath, heart palpitations, chest pain and/or fatigue at ordinary levels of exercise have a NYHA class of II.
3. NYHA Class III corresponds to a marked limitation of physical activity and manifests as shortness of breath, heart palpitations, chest pain and/or fatigue at less than ordinary levels of physical activity.
4. The final NYHA class, IV, refers to patients with symptoms of shortness of breath, heart palpitations, chest pain and/or fatigue that are present even at rest and worsen with physical activity. (30)

Another approach frequently used in the literature is characterizing HF based on the relative output of the left ventricle. Healthy individuals eject approximately 70% percent of the total blood volume contained in the left ventricle during a contraction, and the volume ejected is referred to as the “left ventricular ejection fraction” (LVEF). LVEF is often used as a measure of heart function and is commonly reduced in HF patients. Depending on the level of reduction, HF patients are commonly classified into one of three groups:

1. Heart failure with preserved ejection fraction (HFpEF) – LVEF $\geq$50%.
2. Heart failure with mildly reduced ejection fraction (HFmrEF) – LVEF from 41 to 49% (previously called heart failure with midrange ejection fraction).
3. Heart failure with reduced ejection fraction (HFrEF) – LVEF $\leq$ 40%. (1)

Along with the NYHA classification, classification according to ejection fraction is frequently used in clinical guidelines. Historically, HFrEF has been the major focus of research and treatment guidelines, but contemporary research is increasingly concerned with HFpEF, which some studies indicate might be much more prevalent than previously suspected and might surpass the prevalence of HFrEF (31). For the purposes of this thesis, the focus will be on HFrEF, as the scientific knowledge base on HFpEF is still maturing.

In summary, current international guidelines, defines the diagnosis and classification of HF as reliant on two factors, the observed symptoms of the patient and objective measurements of cardiac output, such as the LVEF measured by echocardiography (1,32). A positive finding of either cardiac symptoms with no other explanation and/or reduced LVEF results in an HF diagnosis. Local guidelines might differ slightly from the guideline recommendation; for example, the Danish guidelines for HF diagnosis require objective measurements of cardiac dysfunction (33). This complicates the diagnosis of HFpEF, but also reduces the risk of misclassifying another disease as HF (32).

2.3. EPIDEMIOLOGY OF CHRONIC HEART FAILURE

Currently, the prevalence of HF is estimated to be 1-2% of the adult population in Western society (1), with a yearly incidence rate of 0.5% for adults (1). HF disproportionally affects elderly individuals, with an estimated prevalence of more than 70% in those 70 years or older, while the prevalence is below 1% for individuals younger than fifty-five. Modern treatments have succeeded in lowering the mortality of the underlying causes of HF, which has resulted in more patients living with HF. While better preventative care and management of cardiovascular disease is expected to reduce the age-adjusted incidence rate in the coming years, the demographic shift in age counteracts this effect, and the overall incidence of HF is projected to increase (1,5). In addition to the increased incidence of HF, better management of HF has lowered the associated mortality, resulting in a high number of individuals living with CHF (2–4).

CHF is associated with increased mortality and morbidity; however, estimations of the exact magnitude are difficult due to differences and misclassification in most registries (7). One meta-analysis estimate puts the 1-, 2-, 5- and 10-year mortality rates at 87%, 73%, 57% and 35%, respectively (9,34), but observes large variances depending on the exact criteria used by the included studies, with the lowest mortality rates observed in studies including patients treated in the secondary sector (34). In Denmark, a registry-based study of all CHF patients found the 1- and 5-year mortality rates to be 33% and 43%, respectively (4). While the mortality rates remain high, it is important to note that they have declined substantially since 2000 (3,4,9).

Individuals with CHF require frequent health care interventions and experience an average of 1 hospitalization per year (9,35). It is estimated that in 1-2% of all hospitalizations, heart failure is the primary diagnosis (36). This poses a large strain on health care budgets, with studies showing that treatments related to heart failure account for more than 2% of the total healthcare budget, and these costs are expected to increase in the future (36).

2.4. SOCIOECONOMIC INEQUITY IN CHRONIC HEART FAILURE

CHF disease is correlated with many different sociodemographic variables. In the previous section, the relationship between CHF and age was presented, which is partly explainable by the pathophysiology of the syndrome – progressive deterioration of the cardiovascular system will, given time, inevitably result in failure to adequately supply oxygen to the body. However, large disparities in both the incidence and prognosis of CHF are observed when stratifying for other sociodemographic variables with no clear link to disease progression (9,13,15,37). To the extent that these disparities are the results of societal factors outside the control of individuals, these disparities are sometimes referred to as health inequities. Inequities in health are omnipresent in modern society, with low sociodemographic status being strongly correlated with both increased incidence and poor prognosis. In the following section, the concept of health inequities and how they relate to CHF will be explored.

In the seminal Whitehall II study, Marmot et al. found evidence of substantive differences in mortality and cardiovascular morbidity between sociodemographic groups (38). The Whitehall II cohort has since been extensively studied, and various characteristics have been proposed as potential causes of the observed disparities. Differences in health-related behavior between sociodemographic groups were initially thought to explain the observed difference. Marmot et al. found that characteristics related to health behavior, such as smoking and diet, explained approximately one-third of the observed differences between high and low socioeconomic status individuals (39). It is evident that differences in health behaviors only account for a part of the observed differences, necessitating new explanatory frameworks.

In addition to smoking and diet behavior, a study by Kjesbu et al. found that nonwestern ethnicity is correlated with lower participation in cardiac rehabilitation intervention and higher levels of discontinuation of medicinal treatment (13). In addition, they found that nonwestern ethnicity was associated with the presence of more cardiac risk factors and unemployment. They also found that a high level of education was associated with a higher chance of achieving a VO2max (the maximum rate of oxygen consumption during physical activity) quality goal when receiving cardiac rehabilitation (13). Likewise, Diaz-toro et al. found that CHF disproportionately affects low-income and poorly educated individuals, with additional evidence that low levels of literacy are associated with a higher degree of nonadherence, low use of preventative care and increased use of nonelective hospital treatments (16).

Many mechanisms on both the individual and societal scale contribute to the link between socioeconomic status and poor health outcomes (17). One of the mechanisms through which inequities can propagate is in access to health care services; it is apparent that financial barriers to treatment can discourage patients from seeking appropriate medical care, which is evident in out-of-pocket healthcare systems (40,41). However, access is also limited by other factors, such as geographical circumstances. Hence, geographical distances are large, and patients might be needed to invest a significant proportion of time and money in attending treatments and regular checkups (42). This is true for geographically large countries with areas of low population density, such as America (42,43), but even in small countries such as Denmark, geographically linked inequities in health can be observed (44). In recent years, several telehealth technologies have been developed, and one of the proposed benefits is the potential to greatly mitigate geographical barriers to healthcare access, which will be further explored in section 2.6 and the first thesis paper.

2.5. MANAGEMENT OF CHRONIC HEART FAILURE

Management of CHF is a multifaceted task involving both patients and carers. Optimizing patient outcomes requires a holistic approach, with considerations of pharmacological and nonpharmacological interventions based on the patient's lifestyle, the underlying disease and health literacy (1). A comprehensive description of all the nuances of CHF treatment is outside the scope of this thesis. The goal of this section is to provide a brief overview of the key aspects of CHF management, with an emphasis on CHF with reduced ejection fraction.

Diagnosis and Follow-up

Several of the pharmacological options for treating CHF require uptitration to the correct dosage and the risk of rehospitalization is increased in the first weeks after diagnosis (45). For this reason, some studies indicate that monitoring by an HF specialist during the uptitration phase, can improve outcomes (46–48), and the initial management will often be anchored in a cardiological specialist unit.

When the patient is assessed to be stable, i.e., no further deterioration of the underlying condition or the associated symptoms is expected, the patient may proceed to follow-up in the primary care sector. Despite the high prevalence of CHF, little research has been conducted on the optimal management strategy for follow-up on CHF patients, and the role of specialists in stable care is unclear but should be consulted in the event of deterioration (1). In a departure from previous versions of the European Society of

Cardiology guidelines, the most recent version recommends the use of telemonitoring for follow-up of CHF patients (1). This technology will be explored in section 2.6.

In the following section, the most important pharmacological treatments for CHF symptoms are presented. While the objectives of this thesis are ultimately concerned with telemonitoring of stable CHF patients, pharmacological agents are presented to better explain the interventions available to health care practitioners.

Pharmacological Management

The goal of the pharmacological management of CHF patients is to reduce mortality, limit further deterioration of the condition and mitigate the experienced symptoms (1).

Three categories of pharmacological agents are recommended in the management of all CHF patients, except for very mild cases (NYHA class 1). The three categories are angiotensin-converting enzyme inhibitors (ACE-Is)/angiotensin receptor-neprilysin inhibitors (ARNIs), β-blockers and mineralocorticoid receptor antagonists (MRAs).

ACE-I/ARNI

ACE-I or, if the patients experience adverse or insufficient effects on ACE-I, ARNI are recommended for all categories of CHF patients, unless otherwise contraindicated (1). These drugs impact the renin-angiotensin-aldosterone system (RAAS) by reducing the conversion of angiontensin I to angiontensin II or by antagonistically binding to angiotensin II receptor type 1. In healthy individuals, the RAAS is important in controlling hemostasis, and activation of the pathway causes peripheral vasoconstriction and fluid retention and pharmacological inhibition of the RAAS, ultimately resulting in reduced fluid retention and lower blood pressure for the patient, lessening the strain on the heart and reducing further deterioration. Treatment with ACE-Is significantly reduces all-cause and cardiovascular mortality in CHF patients (49).

MRA

MRA, also known as aldosterone receptor blockers, similarly reduces RAAS activity by inhibiting activation of aldosterone receptors. The addition of MRAs to ACE-Is or ARNIs results in reduced mortality and hospitalization events for CHF patients, with studies indicating a mortality reduction of almost 30% (50,51). MRAs are indicated for all CHF patients with reduced ejection fraction (1). MRAs also have a small but, importantly, potassium-sparing diuretic effect (52).

β-blockers

β-blockers block activation of β-adrenergic receptors, reducing sympathetic activation of the cardiovascular system. The net result is a reduction in heart rate and reduced activation of the renin-angiotensin pathway. β-Blockers are appropriate for all compensated and stable CHF patients with reduced ejection fraction; however, care should be taken in the initial titration process, as β-blockers might reduce cardiac output in the short term, which can exacerbate existing insufficiencies (1). Despite potential short-term adverse effects, treatment of CHF with β-blockers significantly improves mortality and morbidity, with an estimated 34% reduction in one-year mortality (53,54).

In addition, other pharmaceutical agents are recommended, if the CHF patient remains symptomatic on optimal management on ACE-I/ARNI, MRA and β-blockers. These include Sodium-glucose cotransporter-2 (SGLT-2) inhibitors, diuretics, and digoxin.

SGLT-2

SGLT-2 inhibitors are among the newest drugs recommended for routine management of CHF patients. SGLT-2 inhibitors are recommended for all CHF patients with symptomatic heart failure (NYHA class II-IV), despite treatment with ACE-I/ARNI, MRA and β-blockers, regardless of diabetes status (1). Studies have shown that adding SGLT-2 inhibitors to symptomatic CHF patients receiving otherwise optimal medical therapy reduces hospitalizations and mortality by 25% (55,56).

Diuretics

Diuretics play a vital role in managing the symptoms of CHF. The purpose of diuretics in CHF management is to complement the ACE-I/ARNI and MRA regime in maintaining euvolaemia. Loop diuretica is particularly effective in providing a fast, short-lived increase in diuresis and a corresponding reduction in extracellular fluid volume, allowing for dosage adjustments according to the immediate needs of the patient (52,57). In many cases, adjustment of diuretics can be initiated by the patient in response to perceived symptoms of congestion (57).

Digoxin

For some CHF patients, treatment with the abovementioned therapeutics is insufficient; in those cases, Digoxin might be administered. Treatment with Digoxin should be closely monitored due to the narrow therapeutic interval. It is important to note that while Digoxin has been shown to reduce hospitalization, it is unclear whether it provides any benefit in mortality (1).

Nonpharmacological interventions

Several nonpharmacological interventions aimed at improving the health status of CHF patients exist, with exercise-based rehabilitation being highly recommended for all CHF patients in current guidelines (1).

CHF patients are recommended to participate in low-intensity exercise, which can help alleviate symptoms of heart failure. This is particularly true for obese CHF patients, where low-intensity exercise and weight reduction have been found to further reduce symptoms and improve exercise capacity (1). As a consequence, many treatment offers for CHF patients will include an exercise-based component featuring either structured exercise at a center or at-home exercise instruction or a combination of the two (58).

A systematic review and meta-analysis found conflicting evidence on the long-term mortality benefits of exercise interventions for CHF patients (58,59). The overall pooled effect was insignificant (risk ratio (RR): 1.02, 95% confidence interval (CI): 0.70 – 1.51, but a sensitivity analysis excluding the large HF-ACTION study (60) showed evidence of a significant improvement in long-term mortality (RR: 0.62, 95% CI: 0.39-0.98). However, the study found evidence of significantly increased health-related quality of life outcomes for the patients receiving the exercise-based interventions (standard mean difference: -0.57, 95% CI: -0.83). While exercise-based interventions might not improve survival in CHF patients, they remain a strong recommendation in current guidelines based on their impact on quality of life (1). Recently, studies have found evidence that exercise-based interventions significantly improve a composite heart health measure called cardiorespiratory fitness level, which might provide further insight into which patients can benefit from exercise (31).

2.6. TELEMEDICINE

Telemedicine and telehealth/telehealthcare are umbrella terms covering health interventions that use digital/internet-based technologies to receive and transmit data between patients and carers (61). While telemedicine has increasingly gained traction in recent years (62,63), the advent of the COVID-19 pandemic presented a sudden need to remotely deliver healthcare and kickstarted the adaptation of telehealth technologies (18,64,65). While the pandemic is mostly behind us and remote delivery of health care is no longer a strict necessity, many telehealth technologies remain in service, with clinicians and patients describing them as “accessible, comfortable, safe” (66).

Telehealth technologies span a wide range of use cases, such as videoconferencing used to bridge physical distance between carer and patient, telemonitoring of important health parameters, and interventions aimed at promoting patient self-efficacy through health literacy and patient self-management programs. The primary emphasis of this thesis is on telemonitoring, but many, if not most, telemedicine interventions feature aspects of several use cases. In this thesis, telemedicine and telehealth refer to the technologies at large, while telemonitoring refers to a specific technology that transfers health data from the patient to the clinician.

Treatment and management of chronic illnesses have a particular relationship with telehealth technologies because of the patient’s reliance on regular interactions with the healthcare system. With the increasing prevalence of people living with chronic disease in the Western world (62), telemedicine has been suggested as a potentially cost-saving technology that maintains high quality of care (11,65,67). This is also true for CHF in particular, where novel structured telephone support and telemonitoring interventions have been developed (11). Telemonitoring might significantly improve the frequency of measurement of important CHF markers such as weight and blood pressure, allowing for early detection of deteriorating patients (10,68). In addition, telemedicine can aid patients in managing their own disease, improving self-efficacy (69,70). In the context of CHF, telemedicine might reduce the need for hospitalizations, improve early detection and intervention and reduce disease progression (62). In addition, telemedicine has the potential to reduce geographical inequities in access to health care (42,71).

Studies investigating the effects of telehealth technologies for CHF are somewhat conflicting, with some studies showing little or no health benefit of the intervention compared with usual care (62,72,73), while other studies have found significant improvements in mortality and hospitalizations for groups receiving telemedicine (74). A Cochrane review and meta-analysis found a significant effect of telemonitoring technologies, with reduced all-cause mortality (RR 0.80, 95% CI 0.60

to 0.83) and reduced heart failure-related hospitalizations (RR 0.85, 95% CI 0.77 to 0.93) (11). The current European Society of Cardiology guidelines state that home-telemonitoring using noninvasive technologies may be considered for CHF patient management to reduce mortality and hospitalization risk (1).

Home-telemonitoring using noninvasive technologies is prevalent within interventions aimed at CHF patients (1,11,22,68). They are primarily aimed at the early detection of adverse developments, such as decompensation events. They typically consist of the patient using electronic devices to measure biometrics related to cardiovascular health on a regular basis. The devices used vary, but most interventions include measurements of weight, blood pressure and pulse (68). The collected data are then transmitted to a health care practitioner, who is responsible for acting on any concerning values.

Detection of concerning values has proven to be a difficult endeavor, and most clinicians default to a threshold-based approach using rule-of-thumb (RoT) values such as “weight gain exceeding 2 kg over 2 days” (68,75). However, heuristics such as daily change in weight have limited sensitivity (76,77), which means that clinicians might fail to identify CHF patients at risk of deterioration. To increase the detection of at-risk patients and improve the workflow of health care practitioners assessing telemonitoring data, more complex statistical algorithms have been developed, frequently in machine learning (ML)/artificial intelligence (AI) and decision support frameworks.

2.7. AI AND CLINICAL PREDICTION MODELS

Algorithmic models built using ML techniques have gained significant prominence in modern media and society, with the models frequently being referred to as AI. Culturally impactful models include the chess solving "Deep Blue" (78), the picture generator "DALL-E 2", and recently the very powerful natural language interpreter "CHAT GPT 4.0"(79). Various approaches to incorporating AI into healthcare have been attempted (80,81), and the specific applications vary enormously, including precision medicine, automated image diagnostics, decision support, robotics and much more (63,81).

In the context of heart failure, ML and AI are also gaining traction. ML techniques have been shown to potentially improve diagnostics and patient management and reduce costs, and the increase in research and development of new models promises even greater benefits in the coming years (10,63)

As noted in the previous chapter, the rise of telemonitoring technologies for CHF presents new challenges for the modern healthcare sector. Frequent measurements of important clinical characteristics might lead to early detection of decompensation and better care, but this requires strategies for analyzing and reacting to substantial amounts of patient data. One proposed solution is incorporating ML and AI techniques in the telemonitoring solution, in many cases by building alert or risk prediction models that analyze patterns in the data (68,75,82–84).

Development and validation of clinical prediction models

Most clinical prediction models are deceptively simple in their operation and feed the model data, and the intrinsic algorithm outputs a prediction. While proper implementation of the input and output of the model can be challenging, the real difficulty rests in the development of the algorithm and model. In this section, the methodology of developing clinical prediction models will be presented to provide context to the methods of papers II and III.

Development

The development of the ML-based prediction model can be roughly divided into five steps (adapted and condensed from the seven steps presented in Steyerberg, 2019 (85)):

- Step 1 - Preliminary problem and data definition
- Step 2 - Coding of predictors and outcomes
- Step 3 - Model selection and specification
- Step 4 - Model parameter estimation
- Step 5 - Internal and external performance evaluation

Step 1 - Preliminary problem and data definition

The first step of developing a clinical prediction model is defining the prediction problem – what should the model predict?

In CHF research, the prediction problem will often be related to future health events, and the outcome will often be some measure of CHF deterioration, such as a decompensation event, hospitalization or mortality (82).

When the prediction problem of interest has been decided, the next task is investigating what is already known, which predictors are of particular interest and what data sources are available. The clinical researcher should investigate literature on the problem and note previously identified predictors of the outcome of interest. In the case of CHF, it would be surprising to develop a model not accounting for change in the weight of the patient, as weight has been previously identified as a predictor of future adverse events (21). At the same time, the researcher should be mindful of which data sources, which are or can be made available for the development process, and consider which predictors are available and the quality of the data. Depending on the prediction task at hand, data might originate from administrative health registries, such as the Danish National Patient Registry (86), from internet-connected devices such as those used in telemonitoring interventions, from clinical trials or a host of other possibilities. Once a conclusion is reached on which predictors and data sources should be used, the researcher can move to the second step.

Step 2 - Coding of predictors and outcomes

Depending on the prediction problem and data sources determined in step 1, the data collected might need either minor or major adjustments to be suitable for model development. In this step, the researcher should attempt to make informed decisions on how to operationalize the outcome and predictors from step 1. For each variable, it should be considered whether a categorical, continuous, or transformed approach should be taken. In addition, the strategy for handling missing data should be decided.

The outcome in clinical prediction models generally falls into one of two categories – categorical, such as the presence of a deterioration event or continuous, such as change in EuroQol-5 Dimension (EQ5D) health utility score, a measure of general health which ranges between 0-1. In most studies investigating prediction models for CHF, a categorical outcome is chosen (87).

No easy shortcut exists for how the predictors should be managed, but care should be taken to investigate whether previous studies have successfully operationalized the predictors of interest. One example is how features explored in the second paper of this thesis are used for proper model development in the third.

The model developer also needs to address the issue of missing data. Handling missing data in clinical decision models is often nontrivial due to the potential for data being "missing at random" or "missing not at random" (88).

Missing data can be classified into one of three categories, "missing completely at random", "missing at random" and "missing not at random". Missing completely at random, occurs when a mechanism with no relation to the predictors or outcome, causes the missing value – an example could be an administrative registration error when transcoding a survey (85). "Missing at random" occurs when the mechanism causing the missing value is correlated with other observed predictors in the dataset, e.g., elderly are more likely to miss a daily weight measurement (85). "Missing not at random" occurs when the mechanism is correlated to the true value of the missing value itself, e.g., if patient reported hospitalizations are used as the outcome, a large number of hospitalizations might go unreported (85).

Failure to address "missing completely at random" will result in a reduced sample size but is not likely to bias results. Failure to address "missing at random" or "missing not at random" values, will result in a similar reduction in sample size, but also risks biased models. Luckily, options exist for handling missing data when developing clinical prediction models, such as linear interpolation, where a missing value in a time series is replaced with the average of surrounding value or multiple imputation, where a separate model is built to predict the missing value. Proper identification of the missingness mechanism and the selection of an appropriate approach for handling the missing data, rely on the model developers familiarity with the data sources and the factors generating missing values.(85,89)

When the data have been processed for model development, the result should be a single dataset (table) with a row for each observation and a column for each predictor and the outcome. At this point, the developer should consider splitting the dataframe into two sets, a training set used for model development and a testing set used for performance evaluation in the fourth step, which is the recommended approach when dealing with large datasets (90). However, this approach can be inefficient when dealing with small datasets or sparse outcomes, in which case proper cross-validation is a sufficient alternative (85). If the developer wants to use cross-validation for performance assessment, the data should be split into N subsamples, called folds, where N is often pragmatically chosen based on sample size and the distribution of outcomes. When using cross-validation for performance evaluation, the remaining steps are repeated N times, where each iteration uses a unique dataset constructed by combining N-1 of the folds as the training set and the last fold as the validation set (equivalent to the test set of the splitting method).

Step 3 – Model selection and specification

During the third step, the developer must decide upon which candidate models should be considered. The contemporary focus on ML and AI technologies has greatly increased the number of available model frameworks for clinical prediction models. The models can be roughly divided into classes based on their method for generating a prediction. The frameworks used in the papers of this thesis are linear regression models, tree-based models and the RuleFit algorithm which combines features of both.

Linear regression models

Linear regression models are a subgroup of parametric models (i.e., models that rely on prespecified assumptions about the distribution of data) that are ubiquitous in classical biostatistics. Linear regression models are defined by the expected linear relationship between the outcome of interest and the chosen predictors, and this relationship is shown in equation 1:

$$Regression\ assumption:$$

$$E(Y|X)\ is\ linear\ in\ X_{1} \ldots.. X_P$$

This assumption of linearity in linear regression models is both a strength and weakness for the models; when the assumption is true, the prespecified relationship strengthens the performance of the model compared to the assumption-free models presented later in this section. However, when the assumption is false, the linear regression model can fail to capture and model the relationship between predictor and outcome, resulting in poor performance (91). Linear regression models are almost always included in development studies of clinical prediction models, partly because they are relatively simple to implement and show competitive performance when compared with more advanced models (92,93) and partly due to transparency of the developed algorithm, a feature that will be elaborated in step 5.

Tree-based models

Tree-based models, also referred to as "recursive partitioning", rely on decision tree frameworks to reach their prediction. A single decision tree consists of a series of decision rules, consisting of binary if-then statements based on the observed values of the predictors. This closely resembles the RoT heuristics often used by clinicians in the management of CHF patients, as exemplified by figure 2, which shows a decision tree based on Danish monitoring guidelines for weight change in CHF patients (94).

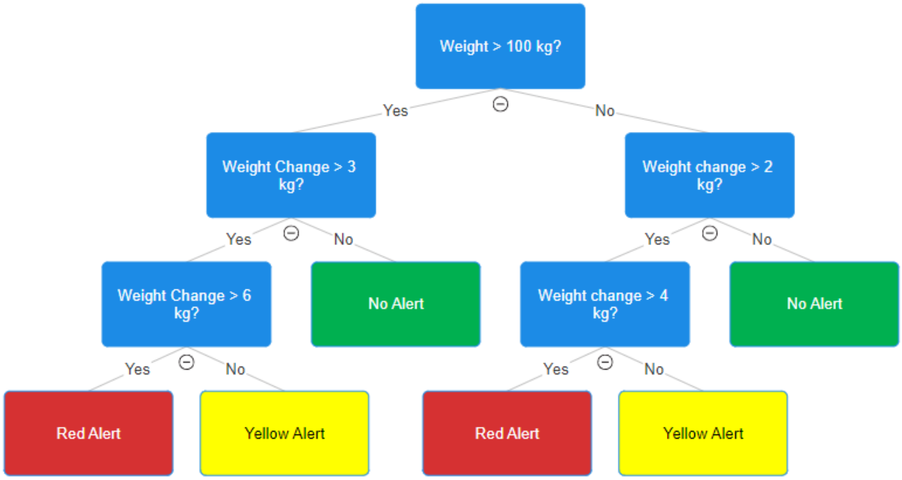


Figure 2 - Decision tree of weight-based thresholds, figure made on smartdraw.com

Like Figure 2, decision trees in clinical prediction models consist of a series of binary splits that result in a number of end points. Each binary split is defined by using the predictor and cutoff value that maximizes between-group variance and minimizes within-group variance with regard to the outcome. The trees are constructed so that each patient belongs to one and only one endpoint, with the majority label (in the case of classification) or the mean value (in the case of regression) of the endpoint representing the patient label. The depth (number of splits) in a decision tree model is defined by the developer, either by a predefined depth or a specified stopping rule, such as minimum number in group at endpoint or minimum improvement in residual variance.

Decision trees are not limited by the assumption of linearity of linear regression models; a decision tree with sufficiently large depth will be able to capture any complex relationships in data, which will result in near perfect prediction for the data used for estimating the splits (training data). However, simple decision trees such as Figure 2 are rarely used on their own in the development of ML models, as they tend to perform poorly when tested on data not used for training, such as new observations (90,95). This phenomenon, known as overfitting, will be further explored in the cross-validation section. This shortcoming can be mitigated when the decision tree framework is extended to include multiple decision trees trained on smaller parcels of the data. One immensely popular algorithm utilizing this method is known as random forest. The random forest algorithm consists of building many decision trees, often hundreds or even thousands, with each tree built on a fraction of the original population and a subsample of the available predictors. The decision of each individual tree is then averaged to produce the outcome.

RuleFit

In the third paper of this thesis, an algorithm known as RuleFit (96) is implemented. The RuleFit algorithm combines aspects of both linear regression models and tree-based models, with the aim of combining the intuitive interpretability of the first with the capacity for modeling complex interactions of the latter.

Development of a RuleFit model involves two steps: the tree-based rule-generating step and the regression step.

The tree-based step involves constructing a large number of shallow decision trees; this can be done using random forest, but often, a slightly more advanced method known as "boosting" is used. It differs from random forest primarily in the way data are sampled for tree construction, with boosting upsampling observations that are erroneously predicted by previous trees (97). Once the trees are generated, the rules are extracted. This process is exemplified in figure 3, which features a reduced version of figure 2.

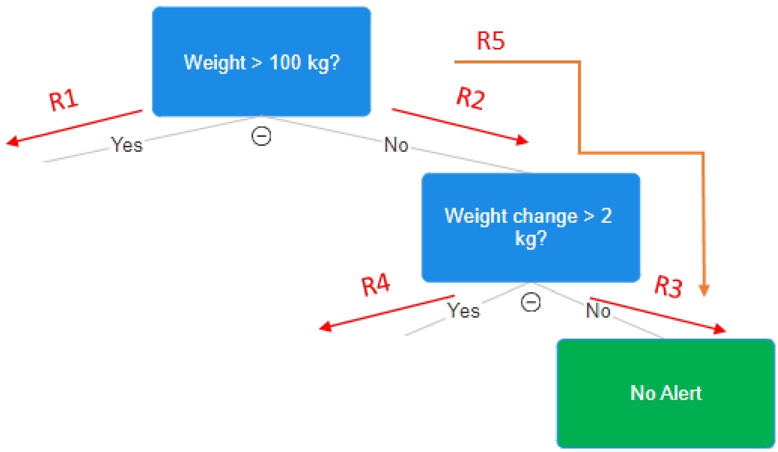


Figure 3 - Example of the rule extraction used in the RuleFit algorithm. Red arrows indicate one condition rule, while the orange arrow illustrates a multicondition rule.

Each splitting rule is individually extracted, as indicated by the red arrows in Figure 3, e.g., the first rule is "weight above 100 kg", as the rule only relies on one condition; this is referred to as a one-condition rule. Multicondition rules are also extracted; in

Figure 3, this is indicated by the orange arrow, which generates the rule "weight below 100 kg AND weight change above 2 kg". When extracting rules, all possible paths through the decision trees are extracted as rules. The trees generated in RuleFit are purposefully kept shallow, which helps limit the complexity and number of rules extracted. Each observation in the original training data is then enhanced with the extracted rules along with binary indicators of whether the rule is true.

In the second step, a linear regression model is trained on the enhanced dataset. As the number of potential predictors has been greatly increased by the first step, it is necessary to use the LASSO regression model, where the penalization term ensures that only the coefficients of the most relevant predictors are nonzero.

The end result of the RuleFit algorithm is indistinguishable from a regular linear regression model, except that the set of predictors included in the model might contain one- or multiple-condition rules.

Step 4 – Parameter estimation

Many of the algorithms used for prediction require the specification of one or more model parameters. Some parameters are best decided by the developer according to the prediction problem, such as the use of logistic versus numerical regression. Other parameters are difficult to optimize based on prior knowledge and require a more data-driven approach; these parameters are sometimes referred to as hyperparameters.

One example is the tree depth of the random forest algorithm. Due to heterogeneous data sources, the optimal depth might differ significantly, even when used for similar prediction tasks, making a priori specification difficult. Luckily, the performance measures elaborated in the fifth step allow for a method of selecting the parameter that leads to the best predictions. The developer must simply build a model for each of the potential values of the parameter of interest and compare performance before selecting the best performing value. Previously, the time and resources needed for this task were prohibitive, but with the efficiency and power of modern computing, searching for a large parameter has become increasingly trivial.

However, one problem arises when using this method for parameter estimation – overfitting. A model is overfit when there is a large decrease in performance between prediction in the data on which the model is trained and prediction on new observations. This occurs when random correlations in the data is interpreted as informative by the model, the risk of this increases with model complexity (a measure of the adaptability of the model, i.e., for tree-based models, increasing the maximum depth increases the complexity).

A simple example: In a dataset on telemonitored CHF patients containing daily measurements and baseline characteristics, all observations of a weight measurement of exactly 79 kg along with a job title of "Artist" occurs prior to a hospitalization, if a tree-based model is trained for this data without any limitations on complexity, it will be likely to predict all new observations of 79 kg artists as "at risk of hospitalization". Developers of clinical prediction models are mostly interested in the validity of predictions for new observations, which means great care should be taken to avoid overfitting (85).

Cross-validation is one method of mitigating the risk of overfitting. Similar to the cross-validation method presented in step 2, the training set is split into N folds (these folds are independent of the folds generated during step 2). A prediction model for each potential parameter value is trained on N-1 of the folds, and the performance is evaluated on the last fold, which is performed N times, with each iteration having a different fold as the test fold. The parameter value with the best performance across the N iterations is then selected for the final model. Once the best hyperparameter values have been selected, the final model can be trained, and the performance can be evaluated.

Step 5 - Performance

The most intuitive way to measure the performance of a prediction model is simply measuring the accuracy on new observations – does the predictions of the model match reality. However, this method of validation is reliant on prospective observations, and few clinicians are willing to use prediction models without a reliable estimate of their expected performance. The data splitting procedure proposed in step 2 and/or cross-validation are alternatives that, under the right circumstances, produce unbiased estimates of future performance of the prediction model (85). Both methods result in a training set, which is used for the training in step 4, and a test set. As the test set has not been included in the model development, it can be considered a reasonable proxy for new observations. When cross-validation is used, the performance estimate is the average of the estimates calculated for each iteration.

All prediction models targeting the clinical sector need to prioritize the identification of all at-risk patients and not give false positives. This prioritization results in accuracy being a poor measure of performance for clinical prediction models, especially when the prediction outcome is rare. In the following section, other, more clinically relevant measures of performance are presented.

Two related terms play an especially important role in understanding the performance of ML algorithms: sensitivity and specificity. Sensitivity is defined as the proportion of true positives marked as positive by the algorithm, i.e., the proportion of patients with imminent deterioration that are accurately labeled as such. Achieving a sensitivity of one, i.e., correctly labeling all at-risk patients as at-risk, is trivial; one only needs to predict the at-risk label in all cases. However, such a prediction would be of little practical use. Thus, sensitivity needs to be balanced by specificity, which is defined as the proportion of true negatives accurately labeled as negative by the algorithm. Balancing the sensitivity and specificity of a model can be achieved by changing the percentage threshold needed for a specific label, i.e., a larger degree of specificity can be achieved by setting a threshold for the positive label above 50%. (85,90)

While sensitivity and specificity are important measures for assessing the performance of a model, summary measures that describe the total discriminatory performance have been developed. The simplest summary measure is the accuracy of the prediction, which is the proportion of correct labels, calculated as the sum of true positives and negatives divided by the total number of observations; however, accuracy is inappropriate when the possible outcomes are not equally likely. Fortunately, this is often the case when dealing with health outcomes, where most people will have a negative label (no disease/deterioration). Most studies reporting on clinical prediction models report the receiver operating characteristic curve (ROC) and the associated area under the curve (AUC). The ROC is generated by plotting pairs of sensitivity and specificity values while varying the threshold from 0 to 1 on a rectangular plot with sensitivity on the y-axis and 1-specificity on the x-axis (illustrated in figure 4). The ROC-AUC is the proportion of the plot beneath the curve.

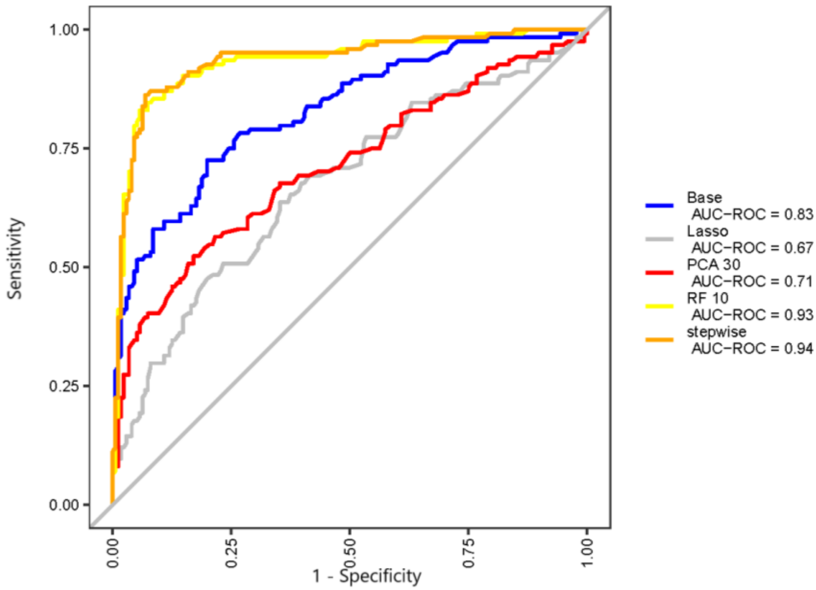


Figure 4 - Example of a ROC plot based on fictional data. Five different predictive models were built, evaluated and plotted, with the best performance belonging to the stepwise model (orange, AUC: 0.94). The R-package "ggplot2"(98) *was used to build the plot.*

A ROC-AUC of 1 indicates perfect prediction, all negatives and positives are labeled correctly by the model, while a ROC-AUC at or below 0.5 indicates no predictive performance of the model (the model is outperformed by random guessing) (85,99). Other performance measures exist, such as precision-recall, f-measure and more, and the most appropriate performance measure depends on factors such as the distribution of outcome labels in the data and the cost of misclassification. The choice of which performance measure(s) to investigate, is ultimately choosing what constitutes a good model. In some scenarios, a good clinical prediction model, might be very conservative in its predictions, resulting in few false positives (high specificity) but also several false negatives (low sensitivity), this could be the case when cost of treatment is very high, while in other scenarios, such as many screening interventions, high sensitivity and low specificity is preferred. In most scenarios, a balanced estimate is preferred, and this makes ROC-AUC the most reported measure.

Prediction models in chronic heart failure

Table 1 presents all identified clinical prediction models for telemonitored CHF patients. The table is based on the work preparing the papers and the findings from three recent reviews, Senarath et al, 2021, which reviewed the literature on prediction models for telemonitored CHF patients to identify the most influential features within the models (21), and Castelyn et al, 2021, which reviewed impact and performance studies of clinical prediction models within multiple chronic diseases, including heart failure (100). The two reviews include both diagnostic and prognostic studies, and while diagnostic models show great promise, they are outside the scope of the current thesis. In the following section, the most important findings from the reviews and the individual studies on prognostic clinical prediction models will be presented to establish the current state of knowledge.

Senarath et al., 2021 found that the most commonly used biometrics were blood pressure, body weight, heart rate and respiratory rate. They found that many prediction models caused a high number of false alerts (low specificity), which further emphasizes the need for a considered approach to balancing sensitivity and specificity in model development. Another important finding was low patient compliance with daily monitoring, to the extent that some studies found evidence of noncompliance being an important predictor on its own. They conclude that future studies on the development of prediction models for CHF patients should focus on incorporating features based on more than one biometric parameter. (21)

Castelyn et al., 2021 also found blood pressure, body weight, heart rate and respiratory rate to be the most commonly measured features, particularly in impact studies, i.e., studies that investigated the health effects of incorporating an algorithm in telemedicine interventions. In addition, they found a large discrepancy between algorithms used in impact studies and algorithms used in performance / development studies, with the former predominantly using rule-thumb-heuristics, while the latter primarily investigated ML-based algorithms (excluding neural nets). They conclude that predictive algorithms are effective in detecting and predicting chronic disease-related events but that closer collaboration between performance and impact studies is necessary. They suggest that performance studies using data and devices available in real-life telemedicine interventions can bridge some of this gap. (100)

While the individual studies presented in Table 1 are grossly heterogeneous, due to differences in the prediction goal, data collected and methodology, some common trends exist. The majority of the studies feature unbalanced datasets, with few events compared to the number of observations (75,101–104).

Some studies developed models that included measurements of intrathoracic bioimpedance (105). However, these models showed conflicting evidence on the importance of the measurements for prediction, and it should be noted that bioimpedance is nontrivial to measure and often requires an implanted device such as a defibrillator (21).

When investigating the potential to improve upon RoT heuristics, two of the studies developed an algorithm based on daily weight measurements, which examined the potential of including long- and short-term averages of weight change in the algorithm compared to using thresholds based on daily weight change (75,101). Zhang et al., 2009 found that models including the "moving average convergence divergence" (MACD) feature had a slightly lower sensitivity but much higher specificity for predicting episodes of worsening heart failure (75). This is partly supported by the findings of Eggerth et al., 2017, who found that the highest specificity belonged to a MACD-based model; however, the best performing models measured by ROC-AUC all featured low specificity (101). Both studies concluded that changes in weight might not be sufficient to identify at-risk patients.

Many of the studies did not report the ROC-AUC for their models (106–110), which makes direct performance comparisons difficult. However, ROC-AUC alone can be misleading in the context of heavily unbalanced data, where a small decrease in specificity might result in numerous false alerts.

In summary, current research suggests that current clinical prediction models for telemonitored CHF patients do not eclipse the performance of RoT-based algorithms. Future development studies should be based on more variables than just weight to increase overall model performance to correct this. Additionally, the models should consider data available in real-life settings, including realistic measurement frequency and compliance, as this constitutes a necessity for later impact and implementation studies.

Study	Data	Model specification	Performance
Eggerth et al., 2017 (101)	Source: HezMobil-Tirol dataset N: 106 patients Observations: 1460 monitoring weeks with 54 events Measured variables: Body weight	Outcome: Heart failure events Algorithm: Comparison between alerts based on MACD and RoT	Aim: Comparison of predictive performance (RoT vs MACD) Performance estimates: MACD AUC: 0.65 Sensitivity 0.87 Specificity0.41 RoT AUC: 0.68 Sensitivity: 0.76 Specificity: 0.57
Gyllensten et al., 2016 (105)	Source: MyHeart Study N: 91 patients Observations: Data divided into 2-week periods (n not reported) Measured variables: Body weight Noninvasive transthoracic bioimpedance	Outcome: Heart failure decompensation Algorithm: Comparison between alerts based on RoT, MACD and Cumulative sums	Aim: Does Noninvasive transthoracic bioimpedance measurements improve performance of previously published weight-based models Performance estimates: RoT AUC not reported Sensitivity: 0.33 Specificity: 0.92

			MACD AUC not reported Sensitivity: 0.50 Specificity: 0.92 Cumulative sums AUC not reported Sensitivity: 0.60 Specificity: 0.96
Henriques et al., 2013 (107)	Source: myHeart telemonitoring database N: 41 patients Observations: 41 two-week episodes with 16 decompensations. Variables Measured: Blood pressure Heart rate Weight Respiration	Outcome: Detection of decompensation episodes Algorithm: Advanced signal processing techniques combined with similarity assessment	Aim: Classification of decompensation episodes Performance estimates: AUC not reported Sensitivity: 0.62 Specificity: 0.69

Henriques et al., 2016 (108)	Source: myHeart telemonitoring database N: 41 patients Observations: 41 two-week episodes with 16 decompensations. Variables Measured: Blood pressure Heart rate Weight Respiration	Outcome: Detection of decompensation episodes Algorithm: Advanced signal processing techniques of multiparametric model combined with case-based reasoning principle Equivalent to k-nearest neighbor method	Aim: Prediction of heart failure decompensation episodes Performance estimates: AUC not reported Sensitivity 0.87 Specificity 0.68
Iakovidis et al., 2016 (109)	Source: Patients using the Motiva telemonitoring system in Kingston-upon-Hull. N: 308 patients Observations: Approximately 10,000 with 3000 events Measured variables: Heart rate Blood pressure Body weight	Outcome: Worsening heart failure Algorithm: Uses signal processing and a Naïve Bayes algorithm	Aim: Worsening heart failure event one day after an *n*-day time-interval of monitoring Performance estimates (4-day monitoring period): Complete Case: AUC not reported Sensitivity 0.738 Specificity 0.965 Linear interpolation: AUC not reported Sensitivity: 0.446, Specificity: 0.966

Javed et al., 2016 (102)	Source: HF patients from London and Dublin N: 115 patients Observations: 735 28- or 21-days blocks (unclear) with 20 events Measured variables: Sound recordings of nocturnal respiration	Outcome: Acute decompensated heart failure Algorithm: Advanced and proprietary signal processing to derive stability estimate combined with alert based on threshold.	Aim: Prediction of acute decompensation Performance estimates: AUC not reported Sensitivity 0.55 Specificity 0.73
Koulaouzidis et al., 2016 (83)	Source: Patients using the Motiva telemonitoring system in Kingston-upon-Hull. N: 308 patients Observations: Data divided into 7,567 (16-day intervals) - 15,761 observations (1-day intervals), labeled as either normal or abnormal periods (n abnormal periods: 13-27) Measured variables:	Outcome: Heart failure hospitalization Algorithm: Advanced signal processing combined with Naïve Bayes	Aim: Detection of patients at high risk of heart failure hospitalization Performance estimates: AUC: 0.82 Sensitivity: 0.52 Specificity: 0.97

	Blood pressure Heart rate Weight		
Ledwidge et al., 2013 (104)	Source: HF patients recruited from St Vincent's University Healthcare Group N: 87 patients Observations: Data divided into just under 2100 weekly periods (exact number not reported) with 1.3% including a clinical deterioration event. Variables Measured: Weight	Outcome: Clinical deterioration Algorithm: Threshold alerts based on MACD (HeartPhone Algorithm)	Aim: Prediction of clinical deterioration Performance estimates: AUC not reported Sensitivity: 0.82 Specificity: 0.68

Zhang et al., 2009 (75)	Source: TEN-HMS study N: 135 patients Observations: 1376 14-day periods classified as either 14-day prior to hospitalization or event free period Measured variables: Daily weight measurements	Outcome: Hospitalization or worsening edema or breathlessness Algorithm: Comparison between alerts based on MACD and RoT heuristics	Aim: Comparison of predictive performance (RoT vs MACD) Performance estimates: MACD AUC = 0.55 Sensitivity: 0.2 Specificity: 0.89 RoT AUC: 0.52-0.53 Sensitivity and specificity not reported.

Table 1 - Summary of previous studies on clinical prediction models for telemonitored CHF patients, compiled from identified reviews and a literature search.

2.8. TELECARE NORD HEART FAILURE

Telemedicine for the management and monitoring of CHF is being introduced across the world, but there are still unanswered questions regarding the potential for mitigating existing geographical inequities in healthcare access. At the same time, the innovative technology promises to improve the possibility of early intervention when the disease inevitably progresses; however, this can only be fully utilized if clinicians are supported by tools and care structures that address the challenges of frequent data measurements.

This thesis is based on data from a Danish telemedicine intervention, "Telecare Nord Heart Failure" (TCNH). In Denmark, efforts are being directed toward incorporating telemedicine technologies in the management of chronic diseases, including HF, at the national level (111,112). A recent randomized trial, the "Telecare Nord Heart Failure trial" (113), has shown promising results. Patients receiving the supplementary intervention had similar physical health and better mental health scores at the 1-year follow-up (22). In addition, the health economic analysis showed supplementary telemonitoring to be a cost-effective alternative, primarily due to reduced hospitalizations in the intervention group (19). This thesis was designed to be a more comprehensive investigation into the data from this trial to potentially improve the targeting and the clinical value of the intervention and interventions like it.

CHAPTER 3. THESIS OBJECTIVES

In this chapter, the overall objective of the thesis will be presented along with the specific objectives investigated by each of the three studies on which this thesis is based.

The objective of this thesis was to investigate whether supplementary telemedicine impacts inequity in health care access for CHF patients. In addition, the potential for predicting CHF deterioration based on weekly measurements of biometric values, with a particular emphasis on investigating the potential of explainable ML models, to ensure patient and clinical acceptance.

The dataset for this thesis originates from telemonitored patients with CHF from the TCNH Trial (22,113).

3.1. SPECIFIC PAPER OBJECTIVES

STUDY	OBJECTIVE
Telemedicine as a Tool for Bridging Geographical Inequity: Insights in Geospatial Interactions from a Study on Chronic Heart Failure Patients	To investigate whether the TeleCare Nord Heart Failure Telemedicine intervention successfully bridged the geographical gap in healthcare access.
Weekly Blood Pressure and Weight Measurements shows Potential for Predicting Deterioration in Telemonitored Chronic Heart Failure Patients	To investigate whether 1-2 weekly measurements are sufficient to predict imminent deterioration in CHF patients.
Prediction of 14-day Hospitalization Risk in Chronic Heart Failure Patients, using Interpretable Machine Learning Methods	To investigate whether prediction models based on interpretable ML algorithms can predict risk of hospitalization within 14 days for CHF patients.

CHAPTER 4. SUMMARY OF PAPERS

4.1. PAPER I - TELEMEDICINE AS A TOOL FOR BRIDGING GEOGRAPHICAL INEQUITY: INSIGHTS IN GEOSPATIAL INTERACTIONS FROM A STUDY ON CHRONIC HEART FAILURE PATIENTS

Purpose: The purpose of this study was to assess the impact of a supplementary telemedicine intervention on geographical inequity in outcomes for CHF patients.

Methods: We utilized a linear regression framework to investigate the relationship between proximity to the usual treatment facility and the benefits of telemedicine. Our study measures changes in health status using EQ5D, 36-item short form health survey (SF-36) physical component score, and SF-36 mental component score as primary outcome measures. The unadjusted analysis included the intervention group, the distance group, and their interaction as independent variables, while the adjusted analysis incorporated multiple socioeconomic and health-related variables to control for potential confounding factors.

Results: Our analysis revealed a significant interaction between the effects of telemedicine and the distance to treatment for changes in EQ5D health status (unadjusted: $p = 0.016$, adjusted: $p = 0.016$) and mental component score (unadjusted: $p = 0.013$, adjusted: $p = 0.008$). However, for the change in the physical component score, the interaction term did not reach significance (unadjusted: $p = 0.118$, adjusted: $p = 0.129$).

Conclusion: Our study suggests that supplementary telemedicine has the potential to mitigate healthcare access disparities associated with geographical distance for CHF patients. Realizing this benefit requires careful implementation of telehealth technologies that prioritize accessibility and adaptation, particularly in underserved areas.

4.2. PAPER II - WEEKLY BLOOD PRESSURE AND WEIGHT MEASUREMENTS SHOW POTENTIAL FOR PREDICTING DETERIORATION IN TELEMONITORED CHRONIC HEART FAILURE PATIENTS

Purpose: The primary objectives of this investigation were to examine the temporal physiological patterns leading up to acute hospital admission events among patients with CHF and to assess the feasibility of utilizing these patterns for predictive purposes.

Methods: Data collected from 33 paired event and control periods, drawn from the TeleCare North Heart Failure trial, served as the foundation for this analysis. The dataset incorporated various features derived from repeated measurements of key physiological indicators such as blood pressure, pulse rate, and weight. These features encompassed statistical measures such as the mean, standard deviation, regression coefficient (slope), difference, absolute change, and max/min ratio. To evaluate their predictive potential in identifying periods preceding deterioration, a logistic regression framework with 5-fold cross-validation was employed.

Results: Weight-pattern alterations manifested in the 30 days leading up to adverse events, while changes in pulse rate and blood pressure occurred closer to the event itself. The three most promising features were identified as the slope of weight between 67 and 37 days prior to an event, the slope of pulse rate during the same period, and the pulse rate difference in the last 37 to 7 days preceding an event. The exploratory model achieved an AUC of 0.81 across the five cross-validation folds, signifying its capacity to predict acute hospital admissions.

Discussion: Our findings suggest that telemonitored data related to weight, pulse rate, and blood pressure in CHF patients can be harnessed to predict acute hospital admissions. While our exploratory model demonstrates strong discriminatory performance in identifying periods at risk of acute hospital admission, it is constrained by a relatively small sample size and its exploratory design.

4.3. PAPER III - PREDICTION OF 14-DAY HOSPITALIZATION RISK IN CHRONIC HEART FAILURE PATIENTS USING INTERPRETABLE MACHINE LEARNING METHODS

Purpose: Our study aimed to assess the predictive potential of biweekly measurements of pulse, blood pressure, and weight for acute hospitalizations in CHF patients. We specifically focused on machine learning models that offer high interpretability, given the limited adoption of complex machine learning models in clinical settings.

Methods: We utilized a dataset comprising 11,575 measurements of pulse, blood pressure, and weight from 122 patients. Three types of machine learning algorithms, namely, logistic regression, random forest, and RuleFit, were trained to predict nonelective hospitalizations within a 14-day window. Performance metrics, including the f-measure, ROC-AUC, sensitivity, and specificity, were estimated using a 5-fold cross-validation framework.

Results: Among the tested machine learning algorithms, a simple and interpretable model, logistic regression with lasso regularization, demonstrated the best performance. The model based on straightforward features achieved an ROC-AUC of 0.622 (sensitivity = 0.185, specificity = 0.93), while the model incorporating a more complex feature set yielded an ROC-AUC of 0.657 (sensitivity = 0.212, specificity = 0.921).

Conclusion: Our study indicates that in the context of predicting hospitalizations in heart failure patients, simple and interpretable methods outperformed more complex black-box machine learning models. This underscores the suitability of interpretable methods in this clinical context. However, it is important to note that the overall model performance was modest, and the limited sample size slightly constrains the strength of our findings.

CHAPTER 5. PAPERS

5.1. PAPER I - TELEMEDICINE AS A TOOL FOR BRIDGING GEOGRAPHICAL INEQUITY: INSIGHTS IN GEOSPATIAL INTERACTIONS FROM A STUDY ON CHRONIC HEART FAILURE PATIENTS

NOTE: This is the submitted manuscript currently undergoing review at BMC Public Health, https://bmcpublichealth.biomedcentral.com/. Permission for inclusion in this thesis has been granted by the publisher.

Authors: **Alexander Arndt Pasgaard Xylander**[*1], Simon Lebech Cichosz[1], Ole Hejlesen[1], Flemming Witt Udsen[1]

[1] Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[*]Correspondence:

Alexander Arndt Pasgaard Xylander

Email: aap@hst.aau.dk

OrcID: 0000-0001-5798-2672

5.1.1. ABSTRACT

Introduction: Chronic heart failure patients experience large disparities in quality of and access to treatment, with rural populations receiving lower levels of care. Telemonitoring of patients is increasingly being used as an important tool for improving patient management and care and might reduce geographical inequities in healthcare.

Methods: We investigate the presence and magnitude of a geospatial interaction effect on the health benefit of a supplementary telemedicine intervention, by analyzing the relationship between distance to regular place of treatment and the benefit of telemedicine in a regression framework. We use change in EQ5D health status, SF-36 Physical component score and SF-36 Mental component score as the outcomes. In the unadjusted analysis, intervention group and distance group and the interaction term are included as the independent variables, in the adjusted analysis, multiple socioeconomic and health related variables are included to account for potential confounders.

Results: We find evidence of a significant interaction between the effects of telemedicine and long distance to treatment for change in EQ5D health status (unadjusted: $p = 0.016$, adjusted $p = 0.016$) and mental component score (unadjusted: $p = 0.013$, adjusted $p = 0.008$), for the change in physical component score the interaction term was not significant (unadjusted: $p = 0.118$, adjusted $p = 0.129$).

Conclusion: In our study we find that supplementary telemedicine is likely to reduce the health access inequities associated with geographical distance for chronic heart failure patients. However, this benefit can only be realized if care is taking to implement telehealth technologies, which emphasizes ease of access and adaptation in underserved areas.

5.1.2. BACKGROUND

Geographical inequity in access to and benefit of health care services has always been a challenge for modern health care systems. The cause of inequity is multifactorial, with complex socioeconomic and cultural factors contributing to disparities in health care. Part of this relationship is the practical barrier caused by distance to the treatment center, and routine care visits might require a significant investment of time and money for a rural resident (42). This is well documented in an American context (42,43), but geographically linked inequity in health is observed even in geographically small countries, such as Denmark(44).

Chronic heart failure is a disease with significant inequities in both incidence and prognosis (13,16,39). This includes geographical inequities. A study found evidence of geographical clusters of acute myocardial infarctions (44), a common precursor to chronic heart failure (114). Another study found a strong correlation between proximity to the treatment center and the odds of receiving proper cardiac revascularization treatment (115).

With the advance of telemedicine as a potent tool in combating the rising costs of healthcare services for chronic diseases, such as chronic heart failure (64,116). Telemedicine has been suggested as a geographical equalizer, with the potential to decrease the geographical inequities of traditional healthcare services (42,71). This makes intuitive sense, as telemedicine can be delivered independently of the distance between carer and patient. However, many health services are still reliant on physical proximity, such as physical exams and surgery, and there is some doubt regarding whether patients can accurately be assessed in a telemedicine setting (117). Inadvertently, telemedicine might even exacerbate existing inequities (18), and socioeconomic factors such as age, area-level income and race appear to be barriers to access to telemedicine interventions for patients with cardiovascular disease (118).

In a randomized controlled trial, an add-on telemedicine intervention (Telecare Nord Heart Failure) has proven successful in reducing health care costs (19) and improving mental health (22). However, it is unclear whether supplementary telemedicine interventions, such as Telecare Nord Heart Failure, have an impact on the health disparity between urban and rural residents. Few studies have investigated telemedicine outside of the urban setting (11), and it is necessary to investigate whether telemedicine is successful in reducing the health inequity of rural populations.

With this study, we wish to investigate whether supplementary telemedicine is associated with a decrease in geographical health inequity for chronic heart failure patients.

5.1.3. METHODS

Research Design

Our aim is to investigate the presence and magnitude of a geospatial interaction effect on the health benefit of a supplementary telemedicine intervention by analyzing the relationship between distance to a regular place of treatment and the benefit of telemedicine in a regression framework. This study attempts to address the question of whether telemedicine is feasible as an approach to combat geographical inequity in healthcare.

We used data from the Telenord Heart Failure study, where a RCT was conducted with chronic heart failure patients randomized to either usual care or a supplementary telemedicine intervention in addition to normal care. The TeleCare North Heart Failure trial aimed to evaluate the health effects and the economic impact of a supplementary telemedicine solution for chronic heart failure patients with 12 months of follow-up. The telemedicine intervention featured at-home measurements of biometric data, such as pulse, weight, and blood pressure. These measurements were evaluated on a weekly basis by health care practitioners, with the hope that continuous measurements would allow a timelier response to any ongoing health deterioration. The original trial included 299 patients, 145 patients allocated to the supplementary intervention and 154 controls from the Region of Northern Denmark.

Patients were presented with a survey at enrollment and at the 1-year follow-up, including items related to socioeconomic status, general health questionnaires and disease-specific items.

The full details of the RCT and the health and health economic effects of the intervention have been reported elsewhere (19,22,113).

Participants and setting

In this study, we included survey data from 168 patients (84 intervention and 84 controls), who had full follow-up on our primary endpoints. The per protocol follow-up time was 1 year, with patient enrollment starting in August 2016.

We include data from related general health questionnaires (EQ5D5L, SF-36) (119,120)and several items pertaining to socioeconomic and health status described in the variables section.

Patient characteristics are reported in the results section.

Variables

Dependent Variable

A total of three different self-reported measures of health are used as the outcomes in this study.

Change in EQ5D health state utility, measured as the difference in momentaneous health utility between EQ5D-5L survey administered at baseline and at the 1-year follow-up. The Danish EQ5D5L scoring was used to calculate the utility score at both points (121). Scoring of the EQ5D5L questionnaire results in a health utility score between 0 and 1, with 1 indicating perfect health and 0 indicating death. In rare cases, scores less than 0 are seen, indicating a health state worse than death.

Change in Physical Component Score, measured as the difference in physical component summary score between baseline and 1-year follow-up, derived from the SF-36 health questionnaire, using the quality metrics proprietary scoring algorithm (120). Scoring of the physical component score results in a score between 0 and 100, with 100 indicating perfect physical health.

Change in Mental Component Score, measured as the difference in physical component summary score between baseline and 1-year follow-up, derived from the SF-36 health questionnaire, using the quality metrics proprietary scoring algorithm (120). Scoring of the mental component score results in a score between 0 and 100, with 100 indicating perfect physical health.

Each outcome was tested independently throughout the analysis.

Independent Variables

Distance from center of treatment

The main interest of this study is investigating the interaction between having a long distance to a regular place of treatment and the benefit of telemedicine. Distance was measured as the distance by car between the patient's home and the treatment provider responsible for the referral to the telemedicine provider and was subsequently grouped into 2 groups, regular and long distance, with long distance defined as the 4th quartile of distance (>40.2 km). The distance by car was collected through a google maps api using the ggmap package (122).

Intervention group

As the data are derived from an RCT, patients are assigned to either supplementary telemedicine or control (treatment as usual). For the purposes of this study, the analysis was conducted as intention-to-treat, with patients being labeled in accordance with the initial randomization.

Age

In the adjusted analysis, patient age is included. Age is defined as the age of the patient at enrollment in the study, calculated from the birth date and date of completion of the baseline survey. For a single patient, an error resulted in the loss of the original completion date. For this patient, age was defined as the mean of the population.

NYHA class

In the adjusted analysis, patient NYHA class was included. The NYHA class is defined as the NYHA class of the patient at enrollment in the study, based upon self-reported status in the baseline survey.

Comorbidity

In the adjusted analysis, the presence of comorbidity is included. Comorbidity is defined as a binary variable indicating the presence of one or more self-reported comorbidities at enrollment in the study.

Gender

Gender was included in the adjusted analysis. Gender was defined by the civil national registry number of the patient.

Marital status

Self-reported marital status was included in the adjusted analysis, and the available response categories were married/registered partnership, cohabiting, partner without cohabitation, single, divorced, and widowed. The responses were recoded into “In a relationship” and “Single” for the first and last three responses, respectively, before inclusion in the analysis.

Employment status

Self-reported employment status was included in the adjusted analysis, and the available response categories were full-time employment, part-time employment, leave of absence, student, sick leave, retired, and unemployed but looking for employment. Responses were recategorized into “employed” for responses 1 and 2, “unemployed” for responses 3, 4, 5 and 7, and “retired” for response 6.

Social Ladder

Self-reported social/societal status was included in the adjusted analysis. Patients were asked to assess their social status on a scale of 1-10, with 10 indicating the societal top and 1 indicating the lowest position in society.

Weight

Self-reported weight at baseline was included in the adjusted analysis.

Handling of numerical and categorical variables

Numerical (continuous) variables were included in the analysis without further modification.

All categorical variables were dummy coded with 1 indicating the presence of the category for use in the analysis.

Main Analysis

To evaluate the presence and strength of an interaction effect between the benefit of telemedicine and distance to the usual treatment center, we used a linear regression framework with an interaction term. For the unadjusted models, the intervention group, distance quartile and intervention*distance group interaction term were included as independent variables. For the adjusted analysis, age, NYHA class, comorbidity, marital status, employment status, social ladder and baseline weight were included as potential confounders. The unadjusted and adjusted coefficient estimates, confidence intervals and p values are reported in Tables 2 and 3, respectively.

Model Assumptions

To ensure that the assumptions of linear regression were not violated, diagnostic plots of the residual errors of the models were generated with the ggfortify package (123) and inspected. Residual vs fitted plots were evaluated to assess linearity, Q–Q plots were evaluated to assess the distribution of errors, and scale-location plots were evaluated for homoscedasticity. As the unadjusted models contained only categorical predictors, the Levene test of homogeneity of variance was used to evaluate homoscedasticity for these models. The results of the visual inspection of the diagnostic plots are reported in the results section, and the full plots and results of the Levene test are shown in appendix 1.

Interaction Effect Interpretation

Coefficients and p values for individual interaction terms are presented in tables 2 and 3 for the adjusted and unadjusted analyses, respectively. However, p values for individual terms are inappropriate for assessing evidence of interaction when the interaction has more than one term (124). To assess whether the models show evidence of interaction, a likelihood ratio test is performed between models with an interaction term and their respective nested model without, and a p value of <0.05 indicates a significant interaction.

Missing Data

We included only patients with complete follow-up on EQ5D, PCS and MCS health scores for this analysis. A small number of patients (n = 7) had missing values in one or more of the categorical variables used in the adjusted analysis. The missing values were mode imputed for the analysis and reported as NA in Table 1.

Statistical Software

All analyses and data management were conducted in R (R version 4.2.1, (125)) using RStudio IDE (RStudio 2023.06.0+421 "Mountain Hydrangea",(126).

The "Tidyverse" package compilation (version 1.3.2, (127)) was used for data management and related tasks. The "Tidymodels" package was used to specify and implement the regression models (version 1.00, (128)). Additional packages were used for various tasks related to visualization of the data and models (123,129,130).

Ethical disclosure and reporting guidelines

The Telecare Nord Heart Failure project has been assessed by The North Denmark Region Committee on Health Research Ethics (http://www.rn.dk/sundhed/til-sundhedsfaglige-og-samarbejdspartnere/forskning/den-videnskabsetiske-komite-for-region-nordjylland), which concluded that the project did not require ethical approval.

This study is reported in accordance with STROBE guidelines on the reporting of observational studies.

5.1.4. RESULTS

No significant differences in baseline characteristics were observed between the patient group randomized to supplementary telemedicine and the patient group randomized to treatment as usual. For two of the primary endpoints, EQ5D health utility score and the physical component summary score of the SF-36, no significant difference in change in health status was observed. However, the change in the mental component summary score of the SF-36 differed significantly between the two groups, with the intervention groups showing a slight increase of 2.65 points, while a slight decline of -0.99 points was observed in the control group.

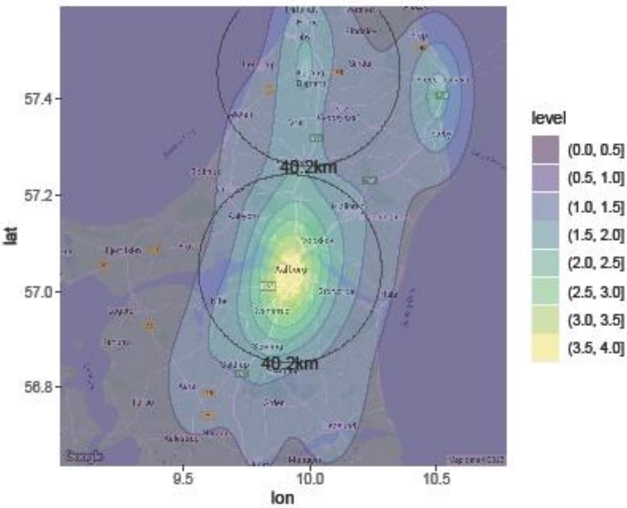
Table of Patient Characteristics			
Variable	Intervention	Control	p value
N	84	84	
Dependent Variables			
Change in EQ5D health utility score (mean (SD))	0.03 (0.19)	-0.01 (0.22)	0.257
Change in SF-36 PCS (mean (SD))	-0.01 (7.75)	1.41 (6.47)	0.200
Change in SF-36 MCS (mean (SD))	2.65 (9.28)	-0.99 (10.25)	**0.017**
Primary Independent variable			
Distance = Long (>40.2 km) (%)	20 (23.8)	19 (22.6)	1.000
Distance in km (mean (SD))	25.41 (20.56)	30.42 (25.80)	0.166

Sociodemographic variables			
Gender = Male (%)	69 (82.1)	67 (79.8)	0.844
Age (mean (SD))	68.99 (9.34)	67.76 (10.32)	0.421
NYHA class (%)			0.912
1	7 (8.3)	6 (7.1)	
2	44 (52.4)	44 (52.4)	
3	24 (28.6)	28 (33.3)	
4	6 (7.1)	4 (4.8)	
Missing	3 (3.6)	2 (2.4)	
Present Comorbidity (%)	46 (54.8)	46 (54.8)	1.000
Weight at baseline (mean (SD))	85.39 (17.15)	86.55 (19.32)	0.681
Marital status (%)			0.059
In a Relationship	65 (77.4)	53 (63.1)	
Single	18 (21.4)	31 (36.9)	
Missing	1 (1.2)	0 (0.0)	
Social ladder score (mean (SD))	5.93 (1.48)	5.99 (1.76)	0.811
Employment status (%)			0.204
Employed	9 (10.7)	16 (19.0)	
Unemployed	9 (10.7)	6 (7.1)	
Retired	64 (76.2)	62 (73.8)	

Missing	2 (2.4)	0 (0.0)	

Paper I Table 1 - Patient characteristics stratified by treatment group

For use in the regression model, distance by car was factored into normal and long distance groups, with the long distance group defined as the fourth quartile (>40.2 km). In Figure 1, the two main treatment centers, located in Aalborg and Hjørring, are shown along with a heatmap of the patient distribution. The solid black circle roughly shows the cutoff between the normal and long-distance groups.



Paper I Figure 1 - Heatmap of the two primary treatment centers with black circles roughly indicating the cutoff point between normal and long distance.

All regression models reported in the following section were evaluated for accordance with the assumptions of a linear regression. No departures from these assumptions were observed in residual vs fitted and QQ plots. The plots are available in appendix 1.

Table 2 shows the results of the unadjusted regression analysis. When examining the effect of long distance on EQ5D, we see that the main effect of long distance is associated with a significant decline in health utility of -0.176 (p = 0.001), and we also see that the coefficient of the interaction term is significant with an increase of 0.178 (p = 0.016). This could indicate that receiving the telemedicine intervention in large parts negates the negative effects of having a long distance to your care provider.

Similar results are seen when examining the change in MCS, where the long distance main effects are associated with a significant -6.509 (p = 0.010) decrease in MCS and the interaction term is associated with a significant 8.804 (p = 0.013) increase in MCS. The pattern is similar when examining PCS, with a main effects estimate of -2.27 (p = 0.224) and an interaction term estimate of 4.09 (p = 0.118); however, the terms do not reach significance.

id	**Term**	**estimate**	**std.error**	**p.value**
EQ5D				
	(Intercept)	0.030	0.025	0.221
	Patient group = Intervention	-0.004	0.035	0.904
	Distance group = Long	-0.176	0.052	**0.001**
	Interaction between Distance = Long and Patient group = intervention	0.178	0.073	**0.016**
PCS				
	(Intercept)	1.92	0.885	**0.031**
	Patient group = Intervention	-2.37	1.256	0.061
	Distance group = Long	-2.27	1.860	0.224
	Interaction between Distance = Long and Patient group = intervention	4.09	2.607	0.118
MCS				
	(Intercept)	0.486	1.19	0.684
	Patient group = Intervention	1.601	1.69	0.346
	Distance group = Long	-6.509	2.51	**0.010**

	Interaction between Distance = Long and Patient group = intervention	8.868	3.51	**0.013**

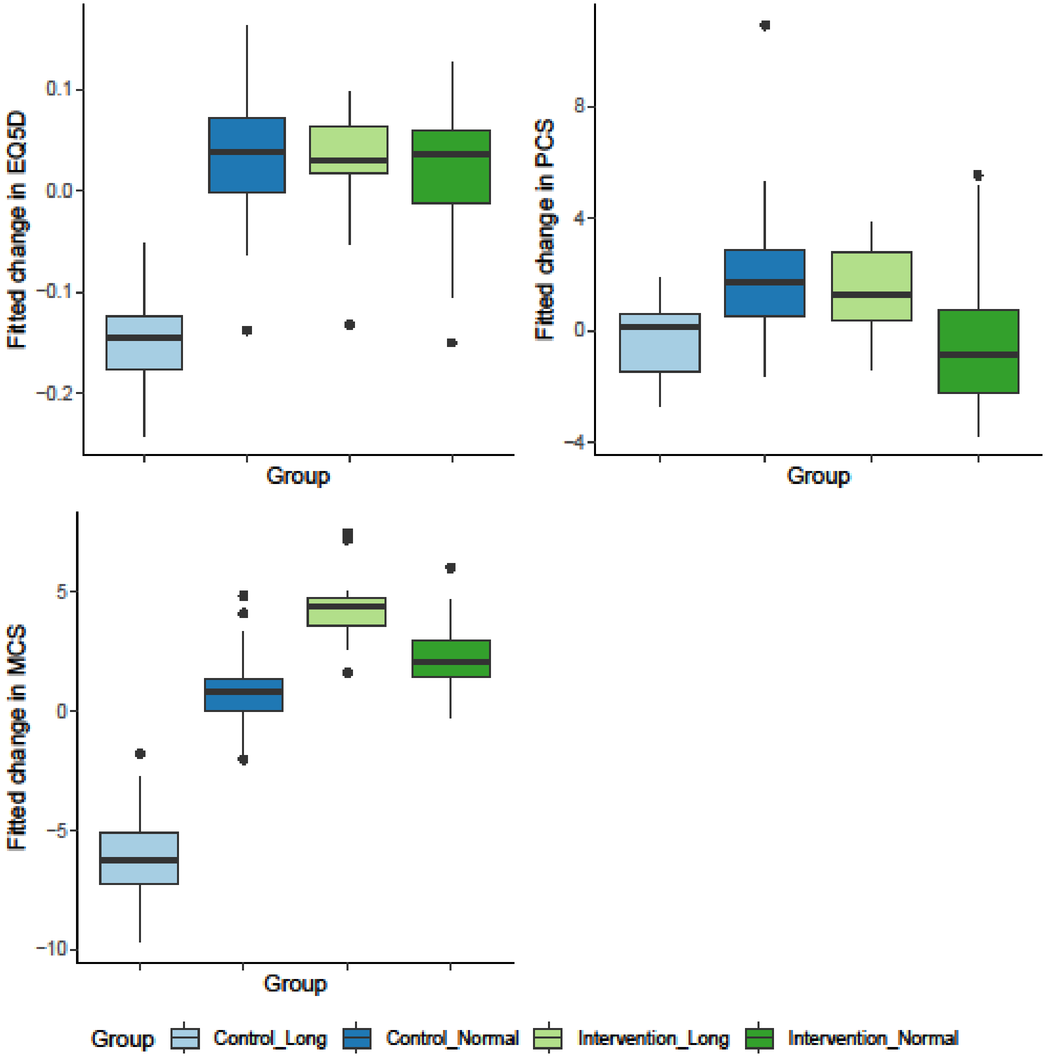
Paper I Table 2 - Coefficients from a linear regression model with a distance-intervention term included. Unadjusted.

Table 3 shows the results of the adjusted regression analysis. We see the same trends as in the unadjusted analysis, with main effect coefficient estimates of -0.197 ($p<0.001$), -2.32 ($p = 0.228$) and -7.67 ($p = 0.004$) for changes in EQ5D, PCS and MCS scores, respectively. The adjusted interaction terms are 0.185 ($p = 0.016$), 4.08 ($p = 0.129$) and 9.79 ($p = 0.008$) for EQ5D, PCS and MCS, respectively. Similar to the unadjusted analysis, the PCS coefficient estimates fail to reach significance. The interaction effects are visualized in Figure 2, which shows a box plot of the fitted values from the adjusted models, stratified by patient and distance group.

id	Term	estimate	std.error	p.value
EQ5D – Adjusted				
	(Intercept)	0.098	0.1171	0.568
	Patient group = Intervention	-0.002	0.037	0.963
	Distance group = Long	-0.197	0.054	**0.000**
	Interaction between Distance = Long and Patient group = intervention	0.185	0.075	**0.016**
PCS – Adjusted				
	(Intercept)	-1.58	6.05	0.794
	Patient group = Intervention	-2.22	1.30	0.090
	Distance group = Long	-2.32	1.92	0.228

	Interaction between Distance = Long and Patient group = intervention	4.08	2.68	0.129
MCS – Adjusted				
	(Intercept)	10.54	8.30	0.206
	Patient group = Intervention	1.30	1.78	0.466
	Distance group = Long	-7.67	2.63	**0.004**
	Interaction between Distance = Long and Patient group = intervention	9.79	3.67	**0.008**

Paper I Table 3 - Coefficients from a linear regression model with a distance-intervention term included. Adjusted for age, Gender, NYHA class, Comorbidity, Weight, Relationship status, Social status, and Employment status.



Paper I Figure 2 - Boxplot of fitted values from the adjusted model, stratified by patient and distance group

Table 4 shows the results of the log-likelihood ratio test. When examining the change in EQ5D health utility, both the unadjusted and adjusted models show significant improvement ($p = 0.014$ and $p = 0.011$, respectively) with the addition of an interaction term between intervention and long distance. This is also true for MCS, with an unadjusted p value of 0.011 and adjusted p value of 0.006. This significant improvement in log-likelihood suggests the presence of long distance effect modification for the general health and mental health effects of the telemedicine intervention. For PCS, the models with interaction terms were not significantly better than the main effects only models, suggesting no evidence of long distance effect modification physical health and the telemedicine intervention.

Term	**LogLik**	**statistic**	**p.value**
EQ5D			
Unadjusted with interaction term	34.6	NA	NA
Unadjusted without interaction term	31.6	5.99	**0.014**
Adjusted with interaction term	38.8	NA	NA
Adjusted without interaction term	35.6	6.46	**0.011**
PCS			
Unadjusted with interaction term	-566	NA	NA
Unadjusted without interaction term	-568	2.5	0.113
Adjusted with interaction term	-553	NA	NA
Adjusted without interaction term	-555	2.54	0.111
MCS			
Unadjusted with interaction term	-617	NA	NA
Unadjusted without interaction term	-620	6.4	**0.011**
Adjusted with interaction term	-606	NA	NA
Adjusted without interaction term	-610	7.64	**0.006**

Paper I Table 4 - Results of the log-likelihood ratio test

5.1.5. DISCUSSION

Our aim was to investigate the relationship between distance to the treatment center and the benefit of telemedicine, measured as changes in the EQ5D-5L health utility score, SF-36 physical component score and SF-36 mental component score. Our results indicate that distance to the treatment center has a significant interaction with the telemedicine intervention when assessing the change in EQ5D and MCS scores. No evidence of interaction was found when examining changes in PCS.

Our results support that supplementary telemedicine might be a potent tool in combating geographical inequities related to distance for chronic heart failure patients. As the intervention examined was supplementary to usual care, our results do not elucidate whether this holds true for telemedicine as a standalone treatment.

While our results show that telemedicine is effective in diminishing an observed negative correlation between long distance to treatment and health scores for chronic heart failure, it is not necessarily clear why this effect occurs. Telemedicine reduces distance as a barrier for seeking treatment (71,117), but other studies have shown that telemedicine might reduce hospitalizations among chronic heart failure patients (19), suggesting that telemedicine might help prevent deterioration of heart disease.

Strengths and weaknesses

Our analysis is consistent across unadjusted and adjusted models, where we have tried to include the most prominent socioeconomic markers, indicating that the observed evidence of significant interaction is unlikely to be an artifact due to residual socioeconomic confounding.

This study is a secondary analysis of data gathered during a previously published RCT (22) with a modest sample size. However, despite the sample size being small, our results are consistent and, in the case of change in EQ5D and MCS, significant across unadjusted and adjusted models, indicating that lack of statistical strength is unlikely to be an issue in our analysis.

The origin of the data means that our analysis risks being impacted by any significant bias in the original RCT. Initial treatment allocation was through randomization, which should minimize any differences in confounders between the patient groups at baseline (124). However, there is a risk of differential dropout between the two treatment arms, with patients experiencing little, no, or even adverse effects from the telemedicine intervention choosing to leave the study. In our complete case analysis, similar dropout rates were observed in the two treatment arms, indicating no difference in dropout. While equal dropout rates in both groups do not preclude nonrandom mechanisms of dropout from causing bias (131), the complete case intervention and complete case control groups were similar at baseline, which indicates that the dropout mechanisms were similar between the groups. While we cannot preclude any residual bias, we believe that our finding of significant interaction is valid, but care should be taken when assessing the strength of the effect, as complete case analysis might bias the effect size if the data do not satisfy the missing completely random conditions (131).

While our analysis is conducted on RCT data and could be representative of experimental conditions, rather than a real-world setting, it is important to note that the intervention arm of the RCT was developed to mirror the subsequent real-world implementation of telemedicine for chronic heart failure patients in the Region of Northern Denmark. This makes it unlikely that our findings are confined to the experimental setting.

In this study, we chose to evaluate three different measures of self-reported health, leading to three main results. We find that the interaction effect of distance to the treatment center is confined to mental and general health and not physical health, and our analysis does not indicate why this is the case. One contributing factor could be specific mental health benefits of telemedicine; a qualitative investigation found that monitoring patient disease through telemedicine can increase the feeling of safety and trust of the patients (66). Chronic heart failure is associated with significantly lower mental health, and while the causality is likely to be multifactorial (132), stressful experiences as a chronic disease patient might play an important role in the relationship, which might be mitigated in part by telemedicine intervention. Future studies should investigate patients' perceived experience of receiving telemedicine, as the increased feelings of safety and comfort associated with telemedicine might be important in preserving the mental health of remote patients.

Comparison with prior work

To our knowledge, this is the first study investigating the interaction effects between distance to a treatment center and receiving a telemedicine intervention. However, previous research has investigated the relationship between digital treatment technologies and socioeconomic factors.

A review by Turnbull et al., 2020 found evidence of socioeconomic factors modifying the effects of web-based self-care interventions for 4 types of chronic illness (133), while our study investigated telemedicine as opposed to a web-based intervention. Some of the potential mechanisms for the propagation of inequities, such as health literacy, technological affinity, and barriers to accessing conventional treatment, might be similar. Turnbull et al. conclude that proper implementation of web-based interventions can prove advantageous in the treatment of traditionally underserved populations.

The recent COVID-19 pandemic has accelerated adaptation and research into telemedicine interventions. One study investigating the differences in the characteristics of patients using telemedicine versus conventional care during the lockdown period found that sociodemographic characteristics differed significantly between groups (134). They found that being younger, married, woman, Hispanic and more factors were all associated with increased odds of receiving telemedicine, which is evidence that the adaptation of telemedicine among vulnerable groups can be low.

Our study provides some evidence that telemedicine can mitigate the effects of geographical barriers to treatment, but care must be taken in the development and evaluation of interventions to ensure their accessibility and adaptation among disadvantaged populations.

Implications

Although distances in a single Region of Denmark are comparatively small, we see a significant interaction between distance and treatment effect. Further research should be conducted on whether the same or perhaps an even stronger association is observed in more rural populations.

Our findings indicate that monitoring patients within a telemedicine framework can be an appropriate tool in delivering high-quality care to remote patients with chronic heart failure. This patient group is a traditionally underserved and vulnerable population, where telemedicine might contribute to reducing health inequities imposed by geographical barriers, particularly barriers related to the distance to treatment.

Conclusion

We found evidence of a statistically significant interaction between the effects of supplementary telemedicine and the patient's travel distance to regular treatment. In our study, belonging to the 4th quartile in distance to treatment center group was associated with decreased general health utility and mental health component scores for the regular treatment group; however, this association was mitigated in the supplementary telemedicine group, indicating that telemedicine counteracts the negative effects of long distance.

This emphasizes that telemedicine is an important tool in providing care for patients located in remote and/or rural areas.

5.1.6. DECLARATIONS

Ethics approval and consent to participate

According to national law in Denmark, all healthcare research involving patient data much be assessed by the regional ethics committee. This study is conducted as a secondary analysis of data collected in the Telecare Nord Heart Failure Randomized Controlled Trial (ClinicalTrials.gov, (NCT02860013)), the trial protocol was assessed by The North Denmark Region Committee on Health Research Ethics (http://www.rn.dk/sundhed/til-sundhedsfaglige- og-samarbejdspartnere/forskning/den-videnskabsetiske- komite-for-region-nordjylland), and the project was assessed as not requiring further ethical approval according to Danish law (135).

Informed consent was obtained for all subjects participating in the study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are not publicly available, as they contain sensitive patient information and access requires approval from the North Denmark Region. Researchers interested in the data may contact the corresponding author for guidance.

Competing interests

The Authors have no competing interests to declare.

Funding

The TeleCare North Heart Failure project and trial was funded by the North Denmark Region, the 11 municipalities in the North Denmark Region, and the Danish Agency for Digital Government. Sponsors of this study did have a role in helping with data collection and project management but did not have a role in conception, analysis, and interpretation of data, writing of the report, or the decision to submit the report for publication.

Authors' contributions

AAPX conceptualized the study, conducted the analysis and wrote the first draft of the manuscript. SLC provided the data and contributed to data management. OH, SLC and FWU contributed to the design of the study, helped interpret the results of the analysis and provided literature and perspectives to the background and discussion sections. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the North Denmark Region, as well as the patients and carers that contributed to original trial.

5.2. PAPER II - WEEKLY BLOOD PRESSURE AND WEIGHT MEASUREMENTS SHOWS POTENTIAL FOR PREDICTING DETERIORATION IN TELEMONITORED

Xylander AAP, Udsen FW, Jensen MH, et al. Weekly Blood Pressure and Weight Measurements shows Potential for Predicting Deterioration in Telemonitored Chronic Heart Failure Patients.

Submitted and currently under review at Journal of Telemedicine and Telecare, 2023

Manuscript not available in the publicly published version of this thesis. The version submitted for consideration by the assessment committee includes the entire manuscript as submitted.

5.3. PAPER III - PREDICTION OF 14-DAY HOSPITALIZATION RISK IN CHRONIC HEART FAILURE PATIENTS, USING INTERPRETABLE MACHINE LEARNING METHODS

Xylander AAP, Cichosz SL, Jensen MH, et al. Prediction of 14-day hospitalization risk in chronic heart failure patients, using interpretable machine learning methods.

Submitted and currently under review at Health and Technology, 2023

Manuscript not available in the publicly published version of this thesis. The version submitted for consideration by the assessment committee includes the entire manuscript as submitted.

CHAPTER 6. DISCUSSION

6.1. SUMMARY OF MAIN FINDINGS

The objective of this thesis was to investigate areas of potential improvement for a supplementary telemonitoring intervention. Partly by investigating whether the intervention is successful in mitigating geographical inequity and partly by investigating the potential for incorporating a clinical prediction model in the intervention to assist clinicians.

Paper I found that belonging to the 4th quartile of distance to the treatment center (the long-distance group) was associated with worse outcomes for the nontelemonitored group. In the telemonitored group, outcomes similar to the normal distance control group were observed in both the long-distance and regular distance groups. When the interaction term was included in a logistic regression model, both the unadjusted model and the model adjusted for socioeconomic factors showed significant interaction between intervention group and distance group for change in EQ5D-5L health utility and change in the mental health component summary score of SF-36. No evidence of interaction was found for the physical health summary score of the SF-36. This suggests that supplementary telemedicine is, at least partly, successful in reducing geographical inequity in CHF management.

Paper II wished to explore whether the data collected during telemonitoring in Telecare Nord Heart Failure showed promise for developing a clinical prediction model for discriminating between stable periods and unstable periods for CHF patients. The best models showed good performance in the curated dataset, and the study found that features based on long-term trends had the greatest impact on model performance.

In paper III, a clinical prediction model to identify at-risk CHF patients based on weekly telemonitoring data was developed based on the findings of paper II. The model was specified with the outcome being hospitalization within the next 14 days. The study found that logistic regression performed better than more advanced ML algorithms, overall model performance (ROC-AUC: 0.657, sensitivity: 0.212, specificity: 0.921) was similar to other studies.

6.2. STRENGTHS AND LIMITATIONS

As with any other scientific work, the methodological choices in the papers presented should be considered when interpreting the results.

The sample

The data used for the papers consist of high-quality data from the Telecare Nord Heart Failure Trial. Data from a telemonitoring implementation study was chosen to ensure that the data mimicked real-life conditions, so a developed clinical prediction model would be ready for implementation. The necessity of ensuring that clinical prediction model development only used data available in actual clinics has been emphasized in recent reviews (21,136). A disadvantage of using these data is the relatively small number of included patients and, for papers II and III, the low rate of hospitalizations.

In frequentist statistical analysis, such as the methods used in paper I, a small sample size impacts the statistical strength of the quantitative analysis, which risks a type II error. However, despite this limitation, the study found significant evidence of interaction in the first paper, suggesting that a larger sample size would have little impact on the findings.

In the development of predictive models, the sample size limits which models and performance estimation procedures are appropriate. For papers II and III, the sample size prohibits the use of a training – test split; a cross-validation setup was used for unbiased performance estimates. The benefit of the cross-validation approach is twofold, when the unbalanced (low number of events) nature of the data is considered: the first benefit is that all data are utilized for estimating the models, and the second benefit is that data splits on low event data might result in the test set having a significantly different distribution of events than the training data (137). While it cannot be precluded that data splitting might have resulted in different performance estimates for papers II and III, there is no indication that the conclusions of the papers would be impacted, as paper II is exploratory, and the performance estimated in paper III is in the range of other studies.

The models

As new ML algorithms are developed at an extremely fast pace, only a small subsection of the available algorithms could be evaluated in paper III. Modern ML algorithms are often complex and are occasionally referred to as "black box" models. A black box model is defined as an opaque algorithm where the specific mechanisms leading to a given prediction cannot be explained (138); this is frequently true for neural networks or even complex tree-based models, such as boosting. In this context, more simple models are considered "interpretable", which is defined as "operations can be understood by a human, either through introspection or through a produced explanation" (139). This is of great importance if clinical decision models are to be implemented in clinical practice, the lack of interpretability has contributed to a surprisingly slow adaptation of clinical prediction models among practitioners (81,140). In addition, some (inter)government regulations, such as the EU's GDPR directive, require interpretability of AI guided decisions in an effort to improve transparency and accountability (140). Adding a clinical prediction model to a telemedicine intervention might also impact the equity of the healthcare service (18). Past models have been known to include racial bias in their predictions, which might exacerbate differentiated treatment outcomes (141). For black box models in particular, this might lead to uncertainties regarding the fairness of the model, which is a major concern in universal healthcare systems that emphasize equal and free access to treatment (141). Based on preexisting literature on the presumed difficulties of clinical adaptation of clinical prediction models (82) and preliminary talks with clinical stakeholders, the thesis papers are based upon interpretable algorithms.

Generalized linear regression models are used in all three papers. Generalized linear regression models are considered interpretable, as by inspecting the coefficients, it is possible to directly assess how each predictor of the model impacts a prediction.

Paper III also included two additional models based upon the popular random forest algorithm and the novel RuleFit algorithm. While the individual trees of a random forest model are easily interpreted, the ensemble method that collects the predictions of hundreds or even thousands of trees precludes interpretation of the total model (96). RuleFit was investigated as a promising alternative that might combine the performance of a random forest model with the interpretability of the linear models.

The first paper was an association study, where a frequentist biostatistical approach to test for an association between an independent variable and a continuous outcome was used. As our tests showed no evidence that the assumptions of linear regression had been violated, a simple linear regression with an interaction term is in line with guideline and textbook recommendations (142).

Our focus on interpretability might have limited performance in the second and third papers. This is of little consequence for paper II, where the performance estimate is exploratory.

paper III found that the prediction models based on logistic regression performed better than more advanced algorithms. This is in line with evidence from reviews on clinical prediction models, which show comparable or better performance from logistic regression when compared with other ML algorithms (93,143). However, this is disputed in a review on clinical prediction models for predicting heart failure readmission and mortality by Shin et al., 2021, which found that ML algorithms outperform regression-based models (144). The three reviews differ slightly in aim and study inclusion criteria, which might explain the different conclusions. Christodoulou et al 2019 included all clinical prediction model studies comparing ML algorithms with logistic regression, regardless of the disease area. Sun. et al, 2022 included all studies featuring clinical prediction models targeting all-cause readmission and/or all-cause mortality for CHF patients and compared the reported performance of studies using conventional statistics with studies using machine learning. The review by Shin et al., 2021 includes studies whose aim is to investigate whether ML outperforms conventional statistics in clinical prediction models for CHF patients. It is possible that more advanced ML algorithms could offer better performance, but this might come at a cost of reduced clinical acceptance and adaptation.

Results

The strong significance of paper I, despite the modest sample size, might partly originate from the quality of the data. The distance variable was computed using high precision estimates of travel distance by incorporating Google Maps data. To the best of our knowledge, this is a novel approach that has not previously been used to study geographical inequities in healthcare access. Previously, studies have relied on area codes or other aggregate measures of location (e.g., (42,44)). However, the novelty also isolates the findings of paper I. As the data originate from a trial, confounding/biasing factors such as differential dropout might affect the observed association between geographical distance and benefit of a telemedicine intervention. Caution is advised before relying on the findings of the paper until other studies can confirm the evidence.

Paper II finds that weekly measurements might be sufficiently informative for generating predictions of hospitalization risk. However, the modest performance of the developed clinical prediction model in Paper III is not in accordance with these findings. While our model performed similarly to other studies, the lack of specificity at clinically relevant levels of sensitivity discourages clinical adaptation of the model.

This discrepancy might be due to the differences in samples between the two papers; paper II included a heavily curated paired sample of 32 event and nonevent periods, while paper III included all available observations, which resulted in a very low event-to-observation ratio.

6.3. IMPACT

The impetus for paper III and in many ways this entire thesis, was the promising results of a similar project within chronic obstructive pulmonary disease (COPD) (145). Larsen et al. showed that COPD exacerbation events could reliably be predicted based on telemonitoring data in addition to population-based parameters (146), and a randomized trial of an algorithm based on these findings has recently concluded, with results pending (147). The goal of any clinical prediction model is real-life implementation; unfortunately, the performance of the models developed in paper III does not match the models developed by Larsen et al., perhaps due to the complex etiology of CHF (82), which precludes direct implementation of our developed algorithm in clinical practice. Based on these findings, it is recommended that further patience is warranted before adapting clinical prediction models for telemonitored CHF patients, in the hope that future clinical prediction models offer a more promising improvement on current heuristics.

Based on the findings of paper I, it is recommended that telemonitoring be considered for all CHF patients living in remote areas. While the findings are based upon one sample and will need to be validated in more settings by future studies, remote patients currently experience significantly decreased access to healthcare services. Telemedicine and/or other e-health solutions might be the only possible alternatives for providing accessible specialist care for these patients, even if the observed reduction in geographical inequity proves erroneous.

6.4. FUTURE RESEARCH

It is difficult to predict whether future clinical prediction models for CHF deterioration can achieve satisfactory performance if other features or model specifications are attempted. Our performance is similar to other contemporary models (82), but future studies might discover useful disease indicators with the potential to greatly improve models.

One of the difficulties associated with the development of clinical prediction models for CHF is the interpatient heterogeneity in etiology and health status. A promising new method for approaching this issue, “mixed effects machine learning” (MEML), burrows from the hierarchical models within classic biostatistics (148,149). In most common ML algorithms, observations are assumed to be “independently and identically distributed,” however, this assumption is often violated in clinical prediction modes, where multiple observations of each patient are frequent. One example of this discrepancy between assumption and reality is seen in Paper III. In MEML models, intrapatient observations are treated as dependent observations by estimating a random effects model. Depending on the method used, this allows for differential slopes and intercepts in the model between patients, which might offer significant performance improvements when the interpatient heterogeneity is large

(85,148). However, MEML models increase in performance with more observations of each patient, suggesting that new patients might need a run-in period before accurate predictions can be made (148).

The impact of telemedicine on healthcare inequities remains uncertain. As presented in and further supported by paper I, one of the assumed benefits of telemedicine is the intrinsic ability to bypass geographical barriers by alleviating the need for physical proximity between patients and carers. However, in paper I, adjusting for other socioeconomic factors does not impact the observed interaction between distance and telemedicine, which strengthens the validity of the findings and indicates that the effects of geographical distance are independent from other socioeconomic factors; thus, removing the geographical barrier will not solve other socioeconomic inequities. Future studies should investigate whether telemedicine interventions for CHF patients impact other aspects of health inequity.

.

CHAPTER 7. CONCLUSION

In this thesis, I have investigated potential avenues for improving telemedicine interventions featuring telemonitoring for CHF patients, qua the impact of geographical barriers to treatment, and the possibility for incorporating clinical prediction models. I have shown that while inequity concerns should still be considered when implementing telemedicine, the intervention can reduce the impact of geographical distance on health outcomes.

The clinical prediction models developed featured modest performance, while the specificity was still better than that of previous RoT heuristics. The lack of sensitivity makes the model inappropriate for clinical adaptation. While the estimated performance is still a slight improvement on current guidelines, the uncertainty associated with validating clinical prediction models dictates that a conservative approach to implementation should be taken.

Data-driven improvements in digital health services hold exciting promise and ominous pitfalls, and I can only hope that my research contributes in a small way to future work within the field.

CHAPTER 8. LITERATURE LIST

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;1–128.

2. Braunwald E. Heart failure. JACC Hear Fail [Internet]. 2013;1(1):1–20. Available from: http://dx.doi.org/10.1016/j.jchf.2012.10.002

3. Christiansen MN, Køber L, Weeke P, Vasan RS, Jeppesen JL, Smith JG, et al. Age-Specific Trends in Incidence, Mortality, and Comorbidities of Heart Failure in Denmark, 1995 to 2012. Circulation. 2017;135(13):1214–23.

4. Schmidt M, Ulrichsen SP, Pedersen L, Bøtker HE, Sørensen HT. Thirty-year trends in heart failure hospitalization and mortality rates and the prognostic impact of co-morbidity: A Danish nationwide cohort study. Eur J Heart Fail. 2016;18(5):490–9.

5. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1789–858.

6. Ho KKL, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: The Framingham Study. J Am Coll Cardiol [Internet]. 1993 Oct;22(4):A6–13. Available from: https://linkinghub.elsevier.com/retrieve/pii/073510979390455A

7. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol. 2016;13(6):368–78.

8. McMurray JJ, Stewart S. Epidemiology, aetiology, and prognosis of heart failure. Heart. 2000;83(5):596–602.

9. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. Eur J Heart Fail. 2020;22(8):1342–56.

10. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: Ready for prime time. Curr Opin Cardiol [Internet]. 2018;33(2):190–5. Available from: https://journals.lww.com/co-cardiology/Fulltext/2018/03000/Machine_learning_in_heart_failure___read y_for.12.aspx

11. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Heart [Internet]. 2017 Feb 15;103(4):255–7. Available from: http://heart.bmj.com/lookup/doi/10.1136/heartjnl-2015-309191

12. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques. Comput Struct Biotechnol J [Internet]. 2017;15:26–47. Available from: https://www.sciencedirect.com/science/article/pii/S2001037016300460

13. Kjesbu I, Prescott E, Hanne Rasmusen HK, Osler M, Larsen ML, Gustafsson I, et al. Socioeconomic and ethnical disparity in coronary heart disease outcomes in Denmark and the effect of cardiac rehabilitation—A nationwide registry study. PLoS One [Internet]. 2022;17(11 November):1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0276768

14. Lawson CA, Zaccardi F, Squire I, Ling S, Davies MJ, Lam CSP, et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: a population-based study. Lancet Public Heal. 2019;4(8):e406–20.

15. Su A, Al'Aref SJ, Beecy AN, Min JK, Karas MG. Clinical and Socioeconomic Predictors of Heart Failure Readmissions: A Review of Contemporary Literature. Mayo Clin Proc [Internet]. 2019;94(7):1304–20. Available from: https://www.sciencedirect.com/science/article/pii/S0025619619300825

16. Díaz-Toro F, Verdejo HE, Castro PF. Socioeconomic Inequalities in Heart Failure. Heart Fail Clin [Internet]. 2015;11(4):507–13. Available from: http://dx.doi.org/10.1016/j.hfc.2015.07.012

17. Elstad JI. Social inequalities in health and their explanations. Oslo Metropolitan University-OsloMet: NOVA; 2000.

18. Haimi M. The tragic paradoxical effect of telemedicine on healthcare disparities- a time for redemption: a narrative review. BMC Med Inform Decis Mak [Internet]. 2023;23(1):1–10. Available from: https://doi.org/10.1186/s12911-023-02194-4

19. Vestergaard AS, Hansen L, Sørensen SS, Jensen MB, Ehlers LH. Is telehealthcare for heart failure patients cost-effective? An economic evaluation alongside the Danish TeleCare North heart failure trial. BMJ Open. 2020;10(1):e031670.

20. Berry C, Murdoch DR, McMurray JJV. Economics of chronic heart failure. Eur J Heart Fail [Internet]. 2001;3(3):283–91. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1016/S1388-9842(01)00123-4

21. Senarath S, Fernie G, Roshan Fekr A. Influential factors in remote monitoring of heart failure patients: A review of the literature and direction for future research. Sensors. 2021;21(11).

22. Cichosz SL, Udsen FW, Hejlesen O. The impact of telehealth care on health-related quality of life of patients with heart failure: Results from the Danish TeleCare North heart failure trial. J Telemed Telecare. 2019;

23. Katz AM. Physiology of the heart. 5th ed. TA - TT -. Philadelphia, PA SE - xv, 576 pages : illustrations ; 24 cm: Wolters Kluwer Health/Lippincott Williams & Wilkins Health Philadelphia, PA; 2011.

24. Ribatti D. William Harvey and the discovery of the circulation of the blood. J Angiogenes Res [Internet]. 2009;1(1):3. Available from: http://vascularcell.com/index.php/vc/article/view/10.1186-2040-2384-1-3

25. Bozkurt B, Coats AJSS, Tsutsui H, Abdelhamid CM, Adamopoulos S, Albert N, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition o. Eur J Heart Fail. 2021;23(3):352–80.

26. McKee PA, Castelli WP, McNamara PM, Kannel WB. The Natural History of Congestive Heart Failure: The Framingham Study. N Engl J Med [Internet]. 1971 Dec 23;285(26):1441–6. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM197112232852601

27. Tsutsui H, Isobe M, Ito H, Okumura K, Ono M, Kitakaze M, et al. JCS 2017/JHFS 2017 guideline on diagnosis and treatment of acute and chronic heart failure — digest version —. Circ J. 2019;83(10):2084–184.

28. Watson RDS. ABC of heart failure: Clinical features and complications. BMJ [Internet]. 2000 Jan 22;320(7229):236–9. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.320.7229.236

29. Metrics GH. Heart failure — Level 1 impairment. 2019;99–100.

30. Dolgin M, Committee NYHAC. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. 9th ed. A Little, Brown handbook TA

- TT -. Boston SE -: Little, Brown Boston; 1994.

31. Omar W, Pandey A, Haykowsky MJ, Berry JD, Lavie CJ. The Evolving Role of Cardiorespiratory Fitness and Exercise in Prevention and Management of Heart Failure. Curr Heart Fail Rep. 2018;15(2):75–80.

32. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Vol. 145, Circulation. 2022. 895–1032 p.

33. Kjærgaard J, Christensen B, Pedersen BK, Jespersen NZ. sundhed.dk. 2021 [cited 2023 Nov 2]. Lægehåndbogen - Hjertesvigt. Available from: https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/hjerte-kar/tilstande-og-sygdomme/hjertesvigt/hjertesvigt/

34. Jones NR, Roalfe AK, Adoki I, Hobbs FDR, Taylor CJ. Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis. Eur J Heart Fail. 2019;21(11):1306–25.

35. Dunlay SM, Redfield MM, Weston SA, Therneau TM, Hall Long K, Shah ND, et al. Hospitalizations After Heart Failure Diagnosis. A Community Perspective. J Am Coll Cardiol. 2009;54(18):1695–702.

36. Shafie AA, Tan YP, Ng CH. Systematic review of economic burden of heart failure. Heart Fail Rev. 2018;23(1):131–45.

37. Verma AK, Schulte PJ, Bittner V, Keteyian SJ, Fleg JL, Piña IL, et al. Socioeconomic and partner status in chronic heart failure: Relationship to exercise capacity, quality of life, and clinical outcomes. Am Heart J [Internet]. 2017;183:54–61. Available from: https://www.sciencedirect.com/science/article/pii/S0002870316302289

38. Marmot MG, Stansfeld S, Patel C, North F, Head J, White I, et al. Health inequalities among British civil servants: the Whitehall II study. Lancet. 1991;337(8754):1387–93.

39. Marmot MG, Shipley MJ, Hemingway H, Head J, Brunner EJ. Biological and behavioural explanations of social inequalities in coronary heart disease: The Whitehall II study. Diabetologia. 2008;51(11):1980–8.

40. Heidenreich PA, Fonarow GC, Opsha Y, Sandhu AT, Sweitzer NK, Warraich HJ, et al. Economic Issues in Heart Failure in the United States. J Card Fail.

2022;28(3):453–66.

41. Gourzoulidis G, Kourlaba G, Stafylas P, Giamouzis G, Parissis J, Maniadakis N. Association between copayment , medication adherence and outcomes in the management of patients with diabetes and heart failure. Health Policy (New York) [Internet]. 2017;121(4):363–77. Available from: http://dx.doi.org/10.1016/j.healthpol.2017.02.008

42. Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: A geospatial assessment. J Am Med Informatics Assoc. 2019;26(8–9):796–805.

43. Laditka JN, Laditka SB, Probst JC. Health care access in rural areas: Evidence that hospitalization for ambulatory care-sensitive conditions in the United States may increase with the level of rurality. Health Place [Internet]. 2009 Sep;15(3):761–70. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1353829209000057

44. Kjærulff TM, Ersbøll AK, Gislason G, Schipperijn J. Geographical clustering of incident acute myocardial infarction in Denmark: A spatial analysis approach. Spat Spatiotemporal Epidemiol. 2016;19:46–59.

45. Lan T, Zhu L, Fan D ming. Mortality and Readmission Rates After Heart Failure : A Systematic Review and Meta-Analysis. 2021;(December).

46. Lund LH, Braunschweig F, Benson L, Ståhlberg M, Dahlström U, Linde C. Association between demographic , organizational , clinical , and socio-economic characteristics and underutilization of cardiac resynchronization therapy : results from the Swedish Heart Failure Registry. :270–9.

47. Savarese G, Carrero J jesus, Pitt B, Anker SD, Rosano GMC, Dahlström U, et al. Factors associated with underuse of mineralocorticoid receptor antagonists in heart failure with reduced ejection fraction : an analysis of 11 2 1 5 patients from the Swedish Heart Failure Registry. :326–34.

48. Lund LH, Carrero J jesus, Farahmand B, Henriksson KM, Jonsson Å, Jernberg T, et al. Association between enrolment in a heart failure quality registry and subsequent mortality — a nationwide cohort study. 2006;

49. Tai C, Gan T, Zou L, Sun Y, Zhang Y, Chen W, et al. Effect of angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers on cardiovascular events in patients with heart failure: A meta-analysis of randomized controlled trials. BMC Cardiovasc Disord. 2017;17(1):1–12.

50. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, et al. The Effect of Spironolactone on Morbidity and Mortality in Patients with Severe Heart Failure. N Engl J Med [Internet]. 1999 Sep 2;341(10):709–17. Available from: http://journals.lww.com/00132586-200006000-00059

51. Zannad F, McMurray JJV, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, et al. Eplerenone in Patients with Systolic Heart Failure and Mild Symptoms. N Engl J Med [Internet]. 2011 Jan 6;364(1):11–21. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM199909023411001

52. Casu G, Merella P. Diuretic therapy in heart failure - Current approaches. Eur Cardiol Rev . 2015;10(1):42–7.

53. Foody JM, Farrell MH, Krumholz HM. β-Blocker Therapy in Heart Failure. JAMA [Internet]. 2002 Feb 20;287(7):883. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.287.7.883

54. Kubon C, Mistry NB, Grundvold I, Halvorsen S, Kjeldsen SE, Westheim AS. The role of beta-blockers in the treatment of chronic heart failure. Trends Pharmacol Sci [Internet]. 2011;32(4):206–12. Available from: http://dx.doi.org/10.1016/j.tips.2011.01.006

55. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. N Engl J Med [Internet]. 2020 Oct 8;383(15):1413–24. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2022190

56. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med [Internet]. 2019 Nov 21;381(21):1995–2008. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1911303

57. Brann A, Tran H, Greenberg B. Contemporary approach to treating heart failure. Trends Cardiovasc Med [Internet]. 2020;30(8):507–18. Available from: https://doi.org/10.1016/j.tcm.2019.11.011

58. Davies EJ, Moxham T, Rees K, Singh S, Coats AJS, Ebrahim S, et al. Exercise training for systolic heart failure: Cochrane systematic review and meta-analysis. Eur J Heart Fail. 2010;12(7):706–15.

59. Long L, Mordi IR, Bridges C, Sagar VA, Davies EJ, Coats AJS, et al. Exercise-based cardiac rehabilitation for adults with heart failure. Cochrane Database Syst Rev. 2019;2019(1).

60. Whellan DJ, O'Connor CM, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Heart Failure and A Controlled Trial Investigating Outcomes of Exercise TraiNing (HF-ACTION): Design and rationale. Am Heart J. 2007;153(2):201–11.

61. Miller EA. Solving the disjuncture between research and practice: Telehealth trends in the 21st century. Health Policy (New York). 2007;82(2):133–41.

62. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet [Internet]. 2011;378(9792):731–9. Available from: http://dx.doi.org/10.1016/S0140-6736(11)61229-4

63. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: Capabilities, features, barriers, and applications. Sensors Int [Internet]. 2021;2(January):100117. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2666351121000383

64. Silva-Cardoso J, González Juanatey JR, Comin-Colet J, Sousa JM, Cavalheiro A, Moreira E. The Future of Telemedicine in the Management of Heart Failure Patients. Card Fail Rev. 2021;7:2–6.

65. Randall MH, Winchester DE. The New Role of Telehealth in Contemporary Medicine. Curr Cardiol Rep [Internet]. 2022;271–5. Available from: https://doi.org/10.1007/s11886-022-01640-5

66. Ekstedt M, Nordheim ES, Hellström A, Strandberg S, Hagerman H. Patient safety and sense of security when telemonitoring chronic conditions at home: the views of patients and healthcare professionals - a qualitative study. BMC Health Serv Res [Internet]. 2023 Jun 6;23(1):581. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09428-1

67. Lin MH, Yuan WL, Huang TC, Zhang HF, Mai JT, Wang JF. Clinical effectiveness of telemedicine for chronic heart failure: A systematic review and meta-analysis. J Investig Med. 2017;65(5):899–911.

68. Brons M, Koudstaal S, Asselbergs FW. Algorithms used in telemonitoring programmes for patients with chronic heart failure: A systematic review. Eur J Cardiovasc Nurs [Internet]. 2018;17(7):580–8. Available from: https://journals.sagepub.com/doi/abs/10.1177/1474515118786838

69. Mizukawa M, Moriyama M, Yamamoto H, Rahman M, Naka M, Kitagawa T, et al. Nurse-led collaborative management using telemonitoring improves quality of life and prevention of rehospitalization in patients with heart failure

a pilot study. Int Heart J [Internet]. 2019;60(6):1293–302. Available from: https://www.jstage.jst.go.jp/article/ihj/advpub/0/advpub_19-313/_article/-char/ja/

70. Middlemass JB, Vos J, Niroshan Siriwardena A. Perceptions on use of home telemonitoring in patients with long term conditions - Concordance with the Health Information Technology Acceptance Model: A qualitative collective case study. BMC Med Inform Decis Mak [Internet]. 2017;17(1). Available from: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0486-5

71. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JGF. 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 Coc. Eur J Heart Fail [Internet]. 2011;13(9):1028–40. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfr039

72. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in Patients with Heart Failure. N Engl J Med [Internet]. 2010 Dec 9;363(24):2301–9. Available from: http://www.nejm.org/doi/10.1056/NEJMx130045

73. 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. Circulation [Internet]. 2011 May 3;123(17):1873–80. Available from: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.111.018473

74. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan BA, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet [Internet]. 2018 Sep;392(10152):1047–57. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673618318804

75. Zhang J, Goode KM, Cuddihy PE, Cleland JGF. Predicting hospitalization due to worsening heart failure using daily weight measurement: Analysis of the Trans-European Network-Home-Care Management System (TEN-HMS) study. Eur J Heart Fail [Internet]. 2009;11(4):420–7. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfp033

76. Lewin J, Ledwidge M, O'Loughlin C, McNally C, McDonald K. Clinical

deterioration in established heart failure: What is the value of BNP and weight gain in aiding diagnosis? Eur J Heart Fail [Internet]. 2005;7(6):953–7. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.ejheart.2005.06.003

77. Abraham WT, Compton S, Haas G, Foreman B, Canby RC, Fishel R, et al. Intrathoracic Impedance vs Daily Weight Monitoring for Predicting Worsening Heart Failure Events: Results of the Fluid Accumulation Status Trial (FAST). Congest Hear Fail [Internet]. 2011;17(2):51–5. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-7133.2011.00220.x

78. Campbell M, Hoane AJ, Hsu F hsiung. Deep Blue. Artif Intell [Internet]. 2002 Jan;134(1–2):57–83. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0004370201001291

79. OpenAI. GPT-4 Technical Report. 2023.

80. Deo RC. Machine Learning in Medicine. 2015;1920–30.

81. Davenport T, Kalakota R. The Potential for Artificial Intelligence in Healthcare. Futur Healthc J 2019. 2019;6(2):94–8.

82. Mpanya D, Celik T, Klug E, Ntsinjana H. Predicting mortality and hospitalization in heart failure using machine learning: A systematic literature review. IJC Hear Vasc [Internet]. 2021;34:100773. Available from: https://doi.org/10.1016/j.ijcha.2021.100773

83. Koulaouzidis G, Iakovidis DK, Clark AL. Telemonitoring predicts in advance heart failure admissions. Int J Cardiol [Internet]. 2016;216:78–84. Available from: http://dx.doi.org/10.1016/j.ijcard.2016.04.149

84. Gyllensten C. Monitoring heart failure using noninvasive measurements of thoracic impedance [Internet]. researchgate.net; 2018. Available from: https://pure.tue.nl/ws/files/97474897/20180627_Cuba_Gyllensten.pdf

85. Steyerberg EW. Clinical Prediction Models [Internet]. 2nd ed. Cham: Springer International Publishing; 2019. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-3-030-16399-0

86. Sundhedsdatastyrelsen. Landspatientsregisteret ["The National Patient Registry"] [Internet]. [cited 2023 Nov 11]. Available from: https://ncrr.au.dk/danish-registers/the-national-patient-register

87. Di Tanna GL, Wirtz H, Burrows KL, Globe G. Evaluating risk prediction

models for adults with heart failure: A systematic literature review. PLoS One [Internet]. 2020;15(1). Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0224135

88. Rubin DB. Inference and Missing Data. Biometrika [Internet]. 1976 Dec;63(3):581. Available from: https://www.jstor.org/stable/2335739?origin=crossref

89. Austin PC, White IR, Lee DS, van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol [Internet]. 2021;37(9):1322–31. Available from: https://doi.org/10.1016/j.cjca.2020.11.010

90. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Elements. 2009;1:337–87.

91. Molnar C. Interpretable Machine Learning [Internet]. 2nd ed. 2022. Available from: https://christophm.github.io/interpretable-ml-book

92. Lorenzoni G, Sabato SS, Lanera C, Bottigliengo D, Minto C, Ocagli H, et al. Comparison of Machine Learning Techniques for Prediction of Hospitalization in Heart Failure Patients. J Clin Med [Internet]. 2019 Aug 24;8(9):1298. Available from: https://www.mdpi.com/2077-0383/8/9/1298

93. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Calster B Van. REVIEW A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol [Internet]. 2019;110:12–22. Available from: https://doi.org/10.1016/j.jclinepi.2019.02.004

94. Telecare Nord Hjertesvigt. Alarmgrænser [Internet]. 2016 [cited 2023 Nov 7]. Available from: https://rn.dk/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/TelecareNord/Telemedicin-til-borgere-med-hjertesvigt/Instrukser-Hjertesvigt/Almen-praksis,-kommune,-hospital/Hjertesvigt-Forudindstillede-alarmgraenser-2023-07-11.ashx?la=da

95. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. Springer; 2013. Available from: https://faculty.marshall.usc.edu/gareth-james/ISL/

96. Friedman JH, Popescu BE. Predictive learning via rule ensembles. Ann Appl Stat. 2008;2(3):916–54.

97. Doupe P, Faghmous J, Basu S. Machine Learning for Health Services

Researchers. Value Heal [Internet]. 2019;22(7):808–15. Available from: https://doi.org/10.1016/j.jval.2019.02.012

98. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org

99. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [Internet]. Springer; 2009. (Springer series in statistics). Available from: https://books.google.dk/books?id=eBSgoAEACAAJ

100. Castelyn G, Laranjo L, Schreier G, Gallego B. Predictive performance and impact of algorithms in remote monitoring of chronic conditions: A systematic review and meta-analysis. Int J Med Inform [Internet]. 2021;156:104620. Available from: https://doi.org/10.1016/j.ijmedinf.2021.104620

101. Eggerth A, Modre-Osprian R, Hayn D, Kastner P, Pölzl G, Schreier G. Comparison of body weight trend algorithms for prediction of heart failure related events in home care setting. Stud Health Technol Inform [Internet]. 2017;236:219–26. Available from: https://books.google.com/books?hl=en%5C&lr=%5C&id=2UXiDgAAQBAJ%5C&oi=fnd%5C&pg=PA219%5C&dq=%22heart+failure%22+%22tele%22%7C%22telemedicine%22%7C%22telecare%22%7C%22telemonitoring%22+%22prediction%22%5C&ots=bhltaNzOWG%5C&sig=LRK2tN1FMdnEzCZQcRtK7xIl0SY

102. Javed F, Farrugia S, Colefax M, Schindhelm K. Early warning of acute decompensation in heart failure patients using a noncontact measure of stability index. IEEE Trans Biomed Eng. 2016 Feb 1;63(2):438–48.

103. Koulaouzidis G, Iakovidis DK, Clark AL. Telemonitoring predicts in advance heart failure admissions. Int J Cardiol. 2016 Aug 1;216:78–84.

104. 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? Eur J Heart Fail [Internet]. 2013;15(4):447–55. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfs186

105. Cuba Gyllensten I, 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 Med Informatics [Internet]. 2016;4(1):e3. Available from: https://medinform.jmir.org/2016/1/e3/?utm_source=TrendMD%5C&utm_medium=cpc%5C&utm_campaign=JMIR_TrendMD_0

106. 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 bioimpedance. JMIR Med Informatics. 2016 Jan 1;4(1).

107. Henriques J, Rocha T, Paredes S, de Carvalho P. Telehealth streams reduction based on pattern recognition techniques for events detection and efficient storage in EHR. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2013 [cited 2023 Nov 14]. p. 7488–91. Available from: http://ieeexplore.ieee.org/document/6611290/

108. Henriques J, De Carvalho P, Rocha T, Paredes S, Morais J. Multi-parametric prediction for cardiovascular risk assessment. Stud Health Technol Inform. 2016;224:15–20.

109. Iakovidis DK, Douska D, Barba E, Koulaouzidis G. Wavelet-based signal analysis for heart failure hospitalization prediction. Stud Health Technol Inform [Internet]. 2016;224:21–6. Available from: https://books.google.com/books?hl=en%5C&lr=%5C&id=wgy3DAAAQBAJ%5C&oi=fnd%5C&pg=PA21%5C&dq=%22heart+failure%22+%22tele%22%7C%22telemedicine%22%7C%22telecare%22%7C%22telemonitoring%22+%22prediction%22%5C&ots=LCjyc9_9Go%5C&sig=Z8YctO57jgTb78aWQE22qM5p9Xc

110. Javed F, Farrugia S, Colefax M, Schindhelm K. Early warning of acute decompensation in heart failure patients using a noncontact measure of stability index. IEEE Trans Biomed Eng [Internet]. 2016;63(2):438–48. Available from: https://ieeexplore.ieee.org/abstract/document/7174520/

111. National udbredelse af telemedicin til borgere med hjertesvigt ["National rollout of telemedicine for citizens with heart failure"] [Internet]. 2023 [cited 2023 Nov 6]. Available from: https://sundhedsdatastyrelsen.dk/da/strategier-og-projekter/telemedicin/telemedicin_hjertesvigt

112. National udbredelse af telemedicin ["National rollout of telemedicine"] [Internet]. 2023 [cited 2023 Nov 6]. Available from: https://sundhedsdatastyrelsen.dk/da/strategier-og-projekter/telemedicin

113. Cichosz SL, Ehlers LH, Hejlesen O. Health effectiveness and cost-effectiveness of telehealthcare for heart failure: Study protocol for a randomized controlled trial. Trials. 2016;17(1):4–9.

114. Jenča D, Melenovský V, Stehlik J, Staněk V, Kettner J, Kautzner J, et al. Heart failure after myocardial infarction: incidence and predictors. ESC Hear Fail. 2021;8(1):222–37.

115. Gregory PM, Malka ES, Kostis JB, Wilson AC, Arora JK, Rhoads GG. Impact of geographic proximity to cardiac revascularization services on service utilization. Med Care. 2000;38(1):45–57.

116. Cowie MR. Building the new digital world: Launch of the European Heart Journal-Digital Health. Eur Hear J - Digit Heal. 2020;1(1):3.

117. Iellamo F, Sposato B, Volterrani M. Telemonitoring for the Management of Patients with Heart Failure. Card Fail Rev. 2020;6:2–4.

118. Brown KJ, Mathenge N, Crousillat D, Pagliaro J, Grady C, Katz N, et al. Social determinants of telemedicine utilization in ambulatory cardiovascular patients during the COVID-19 pandemic. Eur Hear J - Digit Heal. 2021;2(2):244–53.

119. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual life Res an Int J Qual life Asp Treat care Rehabil. 2013 Sep;22(7):1717–27.

120. Ware J, Kosinski MA, Dewey J. How to Score Version 2 of the SF-36® Health Survey. Lincoln, Qual Inc. 2000;

121. Jensen CE, Sørensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L Value Set: A Hybrid Model Using cTTO and DCE Data. Appl Health Econ Health Policy [Internet]. 2021;19(4):579–91. Available from: https://doi.org/10.1007/s40258-021-00639-3

122. Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. R J [Internet]. 2013;5(1):144–61. Available from: https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

123. Tang Y, Horikoshi M, Li W. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. R J [Internet]. 2016;8(2):474–85. Available from: https://doi.org/10.32614/RJ-2016-060

124. Kirkwood BR, Sterne JAC. Essential medical statistics [Internet]. 2nd ed. Essentials (Wiley-Blackwell) TA - TT -. Malden, Mass. SE -: Blackwell Science Malden, Mass.; 2003. Available from: https://worldcat.org/title/50684158

125. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2022. Available from: https://www.r-project.org/

126. Posit team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA; 2023. Available from: http://www.posit.co/

127. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the {tidyverse}. J Open Source Softw. 2019;4(43):1686.

128. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. [Internet]. 2020. Available from: https://www.tidymodels.org

129. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics [Internet]. 2020. Available from: https://cran.r-project.org/package=ggplot2

130. Yoshida K, Bartel A. tableone: Create "Table 1" to Describe Baseline Characteristics with or without Propensity Score Weights [Internet]. 2022. Available from: https://cran.r-project.org/package=tableone

131. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ. 2013;346(January):1–7.

132. De Hert M, Detraux J, Vancampfort D. The intriguing relationship between coronary heart disease and mental disorders. Dialogues Clin Neurosci [Internet]. 2018 Mar 31;20(1):31–40. Available from: https://www.tandfonline.com/doi/full/10.31887/DCNS.2018.20.1/mdehert

133. Turnbull S, Cabral C, Hay A, Lucas PJ. Health equity in the effectiveness of web-based health interventions for the self-care of people with chronic health conditions: Systematic review. J Med Internet Res. 2020;22(6).

134. Jaffe DH, Lee L, Huynh S, Haskell TP. Health Inequalities in the Use of Telehealth in the United States in the Lens of COVID-19. Popul Health Manag. 2020;23(5):368–77.

135. Lov om et videnskabsetisk komitésystem og behandling af biomedicinske

forskningsprojekter [Internet]. Indenrigs- og Sundhedsministeriet; 2003 p. LOV nr 402 af 28/05/2003. Available from: https://www.retsinformation.dk/eli/lta/2003/402#FN501

136. Castelyn G, Laranjo L, Schreier G, Gallego B. Predictive performance and impact of algorithms in remote monitoring of chronic conditions: A systematic review and meta-analysis. Int J Med Inform. 2021 Dec 1;156:104620.

137. Steyerberg EW. Validation in prediction research: the waste by data splitting. J Clin Epidemiol [Internet]. 2018;103:131–3. Available from: https://doi.org/10.1016/j.jclinepi.2018.07.010

138. Beisbart C, Räz T. Philosophy of science at sea: Clarifying the interpretability of machine learning. Philos Compass. 2022;17(6):1–11.

139. Biran O, Cotton C. Explanation and Justification in Machine Learning: A Survey. IJCAI 2017 Work Explain Artif Intell [Internet]. 2017;8–13. Available from: https://mycourses.aalto.fi/pluginfile.php/1135223/mod_folder/content/0/Biran Cotton 2017.pdf

140. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning-based prediction models in healthcare. WIREs Data Min Knowl Discov [Internet]. 2020 Sep 29;10(5). Available from: https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1379

141. Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and Mitigation of Racial Bias in Clinical Machine Learning Models: Scoping Review. JMIR Med Informatics. 2022;10(5):1–13.

142. Schneider A, Hommel G, Blettner M. Linear Regression Analysis. Dtsch Arztebl Int [Internet]. 2010 Nov 5;107(44):776–82. Available from: https://www.aerzteblatt.de/10.3238/arztebl.2010.0776

143. Sun Z, Dong W, Shi H, Ma H, Cheng L, Huang Z. Comparing Machine Learning Models and Statistical Models for Predicting Heart Failure Events: A Systematic Review and Meta-Analysis. Front Cardiovasc Med. 2022;9(April):1–9.

144. Shin S, Austin PC, Ross HJ, Abdel-Qadir H, Freitas C, Tomlinson G, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. ESC Hear Fail. 2021;8(1):106–15.

145. Larsen TK. PREDICTING EXACERBATIONS IN PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY by. 2019;(December).

146. Kronborg T, Hangaard S, Cichosz SL, Hejlesen O. A two-layer probabilistic model to predict COPD exacerbations for patients in telehealth. Comput Biol Med [Internet]. 2021;128(September 2020):104108. Available from: https://doi.org/10.1016/j.compbiomed.2020.104108

147. Secher PH, Hangaard S, Kronborg T, Hæsum LKE, Udsen FW, Hejlesen O, et al. Clinical implementation of an algorithm for predicting exacerbations in patients with COPD in telemonitoring: a study protocol for a single-blinded randomized controlled trial. Trials. 2022;23(1):1–10.

148. Ngufor C, Van Houten H, Caffo BS, Shah ND, McCoy RG. Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c. J Biomed Inform [Internet]. 2019;89(May 2018):56–67. Available from: https://doi.org/10.1016/j.jbi.2018.09.001

149. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1).

SUMMARY

Alexander Arndt Pasgaard Xylander, M.Sc. in Public Health, holds a background in health data analysis, particularly in epidemiology and health economics. His research, conducted within the Medical Informatics research group at Aalborg University in collaboration with the Region of Northern Denmark, forms the basis of this current thesis. The research comprises three papers: Paper I uncovers healthcare access disparities for chronic heart failure patients, emphasizing telemedicine's potential. In Paper II, weekly biometric measurements prove promising in distinguishing stable periods from those preluding hospitalization. Paper III develops a predictive model, using interpretable algorithms, forecasting hospitalization risk with a 14-day lead time.

The thesis concludes by stressing the necessity for further breakthroughs before predictive models are clinically adopted. It highlights research directions and underscores telemedicine's positive impact on addressing geographical inequities. Xylander's work reflects dedication to advancing healthcare research and addressing critical challenges in the field.

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-588-4

AALBORG UNIVERSITY PRESS