UNIVERSITY OF TWENTE.

Towards a comprehensive algorithmic decision-support system in public service delivery:

The case of wearable continuous vital signs monitoring in hospitalized patients.

Autor:

Siccha Lazaro, Bertha Estefany

S2958023

[b.e.sicchalazaro@stuent.utwente.nl]

Submitted in partial fulfillment of the requirements for the degree of Master of Science, program Public Administration, University of Twente

Supervisors:

Dr. Caroline Fischer, Assistant Professor UT

Dr.Ir. Gréanne Leeftink, Assistant Professor UT

Year

2023

Abstract

In recent years, decision-support systems have become a useful instrument for boosting efficiency and productivity in the provision of public services. However, the low sensitivity to individual settings and the lack of representation of certain groups might enhance systemic biases in such decision-support systems. The present study seeks to evaluate the accuracy of data used in algorithmic decision-support systems and the degree to which general standards in such systems reflect individual characteristics. Specifically, the study focuses on an algorithmic decision-support system deployed in a healthcare setting and designed to detect early patient deterioration through continuous monitoring of vital signs. The study was conducted at Isala Hospital Zwolle in the Netherlands, utilizing the Healthdot Philips patch to monitor the vital signs of hospitalized patients. The results showed that contextual factors such as activity, patient positioning, and individual factors such as BMI and age, affect the quality of heart rate and respiratory rate data obtained by the device. Similarly, it was observed that individual factors exert an influence on vital sign values. Specifically, heart rate values exhibited a significant increase among younger age groups when compared to other age groups. The results of this study show the importance of critical analysis of the completeness of the dataset used in algorithms for decision-support systems. In addition, the results of this study emphasize the importance of avoiding general approaches to diverse demographic populations.

Table of Contents

1.Introduction

In public management and policy decisions, there has been an increasing use of data-driven decisions in the last few years. Similarly, many decision-support systems have emerged, ranging from basic algorithms such as risk assessment to more sophisticated models that employ artificial intelligence (Chauhan, 2020). The implementation of data initiatives has increased the efficiency of public services by simplifying procedures, accelerating analysis and achieving more accurate predictions (König & Wenzelburger, 2021; Kuziemski & Misuraca, 2020). Even though decision support systems have guided public administration, these systems are not error-free. Concerns are typically raised regarding the operation of the data-driven algorithm, as it may perpetuate biases and discriminatory practices, and thus produce unintended outcomes (Brown et al., 2019).

Therefore, it is essential to recognize that datasets may be biased toward a particular population, and that the algorithm may exacerbate this bias (Kuziemski & Misuraca, 2020). The underrepresentation of marginalized groups can result in decision-makers relying on incomplete data leading to unequal treatment. For instance, the Dutch government's utilization of an algorithm to assess the likelihood of fraudulent childcare benefit claims, inadvertently exposed non-Dutch families to racial profiling, resulting in the ‘childcare benefits scandal’ (Schellevis, 2021). It's worth noting that a significant factor contributing to these challenges is the often poor data quality, one of those stemming from missing data (Kilkenny & Robinson, 2018), thus a comprehensive evaluation of data quality is imperative for fostering diversity and inclusion (Fararouei et al., 2017).

Clinical settings have benefited from the implementation of decision-support systems, with one notable example being the Early Warning System (EWS) score. This system has served as a prognostication tool to identify deterioration in patients and trigger decisions regarding the amount of monitoring or the transferal of the patient (Fang et al., 2020). The EWS is based on detecting vital sign deterioration, which is often followed by adverse events such as cardiac arrests. Vital signs of patients are put into the EWS and depending on predefined thresholds, points are assigned to each vital sign measurement, which are then summed up to provide a final score (Smith et al., 2014). Over time, technological advancements have automated the EWS, transitioning it from a manual assessment to an automatic process, particularly with the integration of wearable sensors in clinical settings (Subbe & Bramley, 2022).

Following the emergence of the COVID-19 pandemic and the strain on healthcare systems worldwide, the adoption of multiple electronic devices has increased, such as remote monitoring options for vital signs (Manta et al., 2020; Santos et al., 2021). These devices enable remote monitoring of vital signs inside and outside the hospital and provide real-time patient data that can be integrated into the EWS. Furthermore, the use of this technology has the potential to enhance patient outcomes by providing more frequent and accurate vital sign measurements (Posthuma, et al., 2020). However, there is limited information available about the differentiation of vital signs using these continuous monitoring devices, because the majority of existing literature has focused on the efficacy of wearable devices (Downey et al., 2018).

Therefore, this study aims to examine the application of algorithmic decision-support system in the delivery of public services through the specific case of continuous vital sign monitoring in a medical setting. The primary objective is to assess the quality of data used for Algorithmic Decision-Support (ADS) systems and understand to what extent do these systems effectively capture and represent individual characteristics. Specifically, the research focuses on the Early Warning Score (EWS), a type of ADS used in healthcare to detect early patient deterioration.

The study will address three distinct sub-questions: (1) What is the quality of vital signals recorded by wearable devices, and is there a correlation between data loss and patient-related factors? (2) How do the heart rate and respiratory rate distributions differ between the most important patient characteristics, such as gender, age and body mass index? (3) what are the implications of the data quality and individual characteristics on decision support systems like the EWS?

To accomplish the research goals, a quantitative approach was used. The Healthdot patch, a wearable device created by Philips Electronics in the Netherlands, was used in the data collection process. A total of 403 hospitalized patients from Isala Hospital-Zwolle were included in this phase of data collection, which lasted from November 2021 to August 2022. A total of 3 mean tests and 1 regression analysis were carried out using the collected data, which included information on patient activity and posture along with other patient characteristics.

The outline of this study is organized into chapters. Chapter 2 includes a literature review starting from the use of ADS's systems to the particular case of the EWS system in healthcare. In Chapter 3 a conceptual framework derived from the academic literature will be presented in a graphical way. Chapter 4 will delve into the methodology employed and the considerations of the data collection process. Chapter 5 will present the findings from the research. Subsequently, Chapter 6 will draw out the discussion based on the research findings in comparison to academic literature, as well as propose future line of research. Lastly, in Chapter 7 the main conclusions of the research will be summarized.

2.Literature Review

2.1 The use of Algorithm Decision-Support systems in the public service delivery

This chapter examines the adoption of Algorithmic Decision-Support (ADS) systems in the public sector, its early implementation, the rationale for its use, the inherent pitfalls and risks of its application and the associated problem with data incompleteness within these technologies. Furthermore, it provides an in-depth exploration of the diverse applications of ADS, with a specific focus on its significance in healthcare. By scrutinizing the evolution and implications of ADS in public service delivery, this chapter aims to stablish a framework about the use of this algorithms into the delivery of essential public services.

2.1.1 The rise of ADS in the public sector

Algorithmic Decision-Support systems have become increasingly prevalent in the public sector since the early 21st century. These systems have been employed to quantify political outcomes, facilitate large-scale decision-making, and address the limitations of human information processing (Zarsky, 2016).

The adoption of ADS systems in the public sector is rooted in the values of efficiency and effective governance. During a time when a new framework known as New Public Management (NPM) emerged as a model for governance, the promise of efficiency in public administration based on private sector management practices became paramount (König & Wenzelburger, 2021). This led to a significant overlap between the core ideas of NPM and the adoption of ADS systems. Both seek to achieve higher standardization and automation, the use of quantifiable indicators, and a greater emphasis on measurable outcomes and performance rather than rigid procedures. As König & Wenzelburger (2021) define, "ADM -Algorithmic Decision Making- systems take NPM-ideas of efficient service delivery through evidence-based decisions to the next level as they use algorithmic systems to implement, or at least inform, such decisions". In this way, ADS systems have been portrayed as helpful tools that generate more efficient decisions compared to human decision-makers.

The initial generations of ADS tools promised quantification and served as the basis for evidence-based policies. These tools were even recommended as one of the primary methods for quantifying outcomes within the policy field (Nilsson et al., 2008). These systems were characterized by explicitly programmed steps that incorporated the knowledge of experts, allowing software agents to draw inferences and reason based on this knowledge (Chauhan, 2020).

The increased availability of data and the development in computational models has led to a Big Data Era. As a result, ADS systems have become increasingly capable of analyzing vast amounts of data which have empowered these systems to provide more accurate and efficient outcomes (Höchtl et al., 2016). By leveraging to techniques such as automated machine learning algorithms, ADS has revolutionized decision-making processes by enabling classification and prediction tasks to be performed with greater precision. Nonetheless, due to the same phenomena, in recent year, the used of ADS systems has been extending to sensitive areas such as policing (Bennett Moses & Chan, 2018; Oswald et al., 2018), childcare and social benefits (Chauhan, 2020; Keddell, 2019; Saxena et al., 2021), and predictive models for criminal recidivism (Miron et al., 2021; Završnik, 2021), shifting the goal of supporting decision-making to a an automated self-learning classification tool.

In conclusion, the rise of ADS in the public sector has been driven by advancements in AI technologies and the availability of extensive datasets. While the initial focus was on quantification and evidence-based policies, ADS has now evolved to encompass diverse areas, including sensitive domains like policing and criminal justice. The shift towards machine learning algorithms has revolutionized decision making, emphasizing classification and prediction. However, as it will be discussed further, careful considerations must be given to the ethical and accountability aspects associated with ADS to mitigate potential biases and risks.

2.1.2 Application and opportunities of ADS in the public sector

One compelling argument that supports the use of Algorithmic Decision Systems is their superior information processing capabilities when compared to human capacities. The use of ADS allowed for the systematic handling of enormous volumes of data, which produced quicker outcomes and increased uniformity in the management of information (Kolkman, 2020). Especially in areas of public Administration referred to as “mass decision making”, such as taxes or social benefits, ADS may result in the development of valuable support systems (Monarcha-Matlak, 2021).

In light of the recognition that human cognitive resources are inherently limited, it becomes evident that ADS offers the potential to assess a wide array of information sources, thereby facilitating improved decision-making. As a result, organizations with more information resources and processing capacity are anticipated to achieve better results (Maciejewski, 2017). By leveraging algorithms, organizations seek to transform data into knowledge in the hope that informed decision-making will lead to improved resource efficiency (Holten Mandøller, et al., 2020).

ADS is frequently viewed as a more objective method than human evaluation because it operates according to a predetermined set of rules. By reducing human biases such as prejudices and stereotypes, the usage of ADS has the potential to improve the objectivity od the decision-making (Kolkman, 2020). The process of human decision-making is frequently impacted by heuristics and biases, which leads to outcomes that are less efficient. This is especially true in situations in which there is a limited amount of information.

The usage of ADS has the potential to provide various benefits, including quicker response times, the analysis of large-scale data sets, and informed decision-making, all of which correspond with the principle of efficiency via enhanced information. However, it is critical to recognize that the fulfillment of these advantages is not assured, because ADS has inherent risks and limitations in public service delivery. As a result, a thorough knowledge of the possible benefits and downsides is required to guarantee the appropriate and successful deployment of ADS systems in the context of public services.

2.1.3 Challenges and risks of using ADS in the public sector

The utilization of algorithmic decision-systems (ADS) presents various challenges and possible risks that require examination, including issues related to transparency, limited contextual comprehension, and biases. Firstly, it is worth noting that numerous algorithms employed in decision-making procedures function as opaque entities, thereby posing challenges in comprehending the fundamental factors and criteria that shape their decision-making processes This phenomenon persists even when employing white-box artificial intelligence , as it requires the involvement of a domain expert capable of conducting audits on these systems(Dourish, 2016).

Moreover, a lack of contextual comprehension presents a substantial peril in the realm of algorithmic decision-support. Algorithms have the risk of overlooking or oversimplifying contextual factors that are essential in decision-making, including social, cultural, and ethical dimensions (Janssen & Kuk, 2016). In healthcare for example, some algorithmics tools oversimplifying the complexity of the patient. This can be seen in the use of the ADM CORONANET, a tool designed to support the admission/discharge decisions of oncology patients with COVID-19. The problem was that the model was unable to provide the logic behind the recommended action and its relationship with the biomarkers (Albumin, C-Reactive Protein, Lymphocytes , Neutrophils and Platelets) levels, especially given the limited number of biomarkers included (Wysocki et al., 2023).

Another important challenge inside the use of ADS is the risk of bias results against marginalized groups (Barocas & Selbst, 2018; Eubanks, 2017; O'Neil, 2016). Even though these biases can emerge at various stages, including data collection, data labeling, algorithm design, and decision implementation, (Janssen & Kuk, 2016) a large amount of these problems arise in the data collection and labeling process (Barocas & Selbst, 2018; Eubanks, 2017; O'Neil, 2016).

In the data collection stage, missing data or absence of demographics recollection can introduce biases (Cahan et al., 2019). Here, the problem arises if in data collection processes systematically exclude or underrepresent certain individuals or communities. For example, studies have shown how the absence of certain variables in datasets such as race or gender, can result in biased outcomes that disproportionately impact marginalized groups (Buolamwini, 2018; Zarsky, 2016).

For instance, an example of bias during the labeling stage[1] can be seen in the case examined by Obermeyer et. al. (2019). The authors scrutinize an algorithm used in US health systems to target high-risk care management. The issue was that the variable “illness" was labeled using “healthcare expenditure by patient" as a proxy, not reflecting actual outcomes. As a result, the algorithm assigns the same risk score to black patients who are much sicker than white patients, but usually spend less on healthcare. Black patients utilize healthcare facilities less for economic or trust-related reasons, producing a bias in the algorithm's forecasts.

[1] Data Labeling is the process of adding labels to provide context to the raw data so that a machine learning model can learn from it. For example, the label can define if an image shows a bird or a car (AWS, 2023)

Another significant danger is that even when utilizing human judgment, some algorithmic biases may still be unintentionally accepted by us. The reasoning behind this is because individuals tend to exhibit a tendency to conform to the decisions made by technology-based system, rather than critically evaluating them (Hitron et al., 2022; Robinette et al., 2017). Hitron et al. (2022) examined conformity biases from a gender-biased robot mediating a male-female argument. Despite being told the robot's algorithm is based on human cases, most participants did not associate the robot's conduct with prejudice. On the contrary, participants justified the robot's actions using explanations related to gender stereotypes. This phenomenon may lead to the belief that these systems can generate comparable biases on a larger magnitude than those resulting from human decisions.

Therefore, it is crucial to acknowledge the risks associated with unreliable data and biases in ADS and to implement measures to minimize their impact. Enhancing data quality through rigorous validation processes, promoting diverse and representative datasets, and incorporating fairness considerations into algorithmic design are vital steps to counteract biases and ensure more equitable decision-making.

2.1.5 ADS applications

Although ADS has numerous uses in the public sector, not all sectors can use it equally. The ability to translate knowledge into ADS systems is more limited in fields like social benefits and caseworkers, where the value of in-person communication and human judgment is stressed (Holten Mandøller et al., 2020). Instead, ADS has demonstrated interesting applicability in the healthcare industry, where decisions involve less discretion and exhibit a more programmatic nature.

For the use in enhanced medical services, ADS has found fertile ground for development in healthcare. Paramount priorities within this sector include the allocation of scarce resources, along with the prediction and prevention of adverse outcomes, and timely diagnoses. In response to these demands, ADS tools have been systematically designed to address such priorities. These tools span various domains including image recognition (Goyal et al., 2020), diagnostic tools (Chowdhury et al., 2020; Kermany et al., 2018), and early warning system (EWS) score to identify risky patients (Capan et al., 2018).

However, the utilization of ADS in healthcare encounters similar challenges as discussed earlier regarding data misrepresentation. Particularly in healthcare, there exists a historical bias in the data collected for medical trials, characterized by imbalances in gender, race, and other factors, leading to non-representative recommendations. For instance, models that helps to develop medicine protocols are based predominantly on European ancestral genotypes (Gijsberts et al., 2015; Paulus et al., 2018).

Although ADS systems are a powerful tool for improving performance and resource effectiveness, they are far from error-free. However, accurate results can be achieved from analyzing and intervening with the causes of biases, thereby lowering potential hazards. In this regard, it is clear that the efficacy of ADS depends on more than just the technology itself, as well as the sources of the data and the guiding principles that guide the development of these datasets and systems.

2.2 The case of the Early Warning System Score in healthcare

The primary focus of this study is to examine a specific application of Algorithmic Decision-Support system in the healthcare sector. The presence of structured healthcare data and the nature of evidence-based approaches, and the evident pressure to counter the scarcity in human resources in healthcare have evidently generated a keen interest in using Big Data and different algorithmics in the healthcare practice. As discussed in the previous section, this context led to an increase in research on assistant chatbots, AI image recognition models, and Machine Learning analysis (Zhao et al., 2021).

However, even prior to this technological revolution, already in 1997 manual algorithms in the form of input-output systems were introduced as tools to assess patient deterioration within hospital settings, utilizing physiological measures such as vital signs (Nagarajah et al., 2022). This marked the introduction of the Early Warning System (EWS), which has since evolved over time. While recent advancements have made it feasible to automate the EWS and enhance ADS systems, concerns persist regarding the accuracy of data captured by these machines and whether conventional thresholds adequately account for individual variations.

The following chapter analyzes these concerns, structured across three sections. The first section presents the vital signs measurements and its relationship with the Early Warning System, the second section talks about the use of wearable devices to continuous monitoring of vital signs and the last section explores the association between individual factors and vital signs thresholds.

2.1.2 Vital signs measurements and the Early Warning System (EWS)

The abnormalities in a patient's vital signs are an important factor in detecting subsequent deterioration. This deterioration can be materialized as what is considered an adverse event such as a cardiac arrest, respiratory arrest or any other complication (Hillman et al., 2002; Sapra et al., 2020). The extent of vital sign abnormalities may also be a predictor of long-term patient health outcomes, a return to the emergency department (ED), and hospital readmissions (Andersen et al., 2016; DeVita et al., 2010). Due to its importance, it is used to assess the patient's level of urgency in the medical setting. For this reason, to account for these shifts in vital signs a number of early warning systems (EWS) have been developed as a method to identify vital sign abnormalities and similar parameters (Williams et al., 2016).

The EWS is used in hospitals as part of a "track-and-trigger" system, in which a rising score results in a stepped-up response, ranging from more frequent patient observations (for a low score) to urgent review by the medical staff (Le Lagadec & Dwyer, 2017). Usually temperature, pulse rate, blood pressure, oxygen saturation, and respiratory rate are the vital signs that are most frequently measured. Diverse adaptations of the Early Warning System have been developed, encompassing supplementary variables and patient-specific specifications.

Noteworthy among these adaptations is the National Early Warning Score (NEWS), which introduces the variable of "patient awareness" (Smith et al., 2013). Another example is the Chronic Respiratory Early Warning System, CREWS, tailored for patients with chronic hypoxaemia , characterized by persistently low oxygen saturation levels even when their condition is stable (Eccles et al., 2014).

However, there are different modifications that include additional variables such as the National Early Warning Score, NEWS, and even different specifications based on the group of patients like the Chronic Respiratory Early Warning System, CREWS, which is adapted for patients with chronic respiratory diseases that often presents low oxygen saturation even when their condition is stable.

The measurement of vital signs has traditionally been performed either partially or entirely by manual means, resulting in potential inaccuracies in measurement and calculation (Alam et al., 2014). In a conventional approach, vital signs are recorded three times daily or at 8-hour intervals. Nevertheless, factors like exhaustion and inadequate personal preparedness can affect the accuracy of these recordings, making it difficult to identify early deterioration by the medical personal (Leuvan & Mitchell, 2008; Petersen, 2018). The effectiveness of the EWS in predicting deterioration has been limited by the challenges associated with manual vital signs assessments (Downey et al., 2017). However, recent developments in wearable sensors offer the potential to replace current manual processes with remote and continuous monitoring of vital parameters, thereby reducing the patient's deterioration (Cardona-Morrell et al., 2016; Weenk et al., 2019).

2.1.3 Evidence of using wearable devices for continuous vital signs monitoring

The potential of incorporating wearable devices in the monitoring of vital signs is showing significant promise. When evaluating the available evidence pertaining to these devices, it is essential to not only evaluate their precision in measuring vital signs but also to scrutinize their ability to detect potential adverse events, including cardiac arrest, respiratory arrest, mortality, and rehospitalization.

In the following subsection, these subjects will be explored to gain a comprehension of the advantages and constraints related to the utilization of wearable devices for patient monitoring within a hospital setting. Through examination of the various aspects, the aim is to offer a comprehensive evaluation of the feasibility and efficacy of wearable devices for continuous monitoring for in-hospital patients.

a. Evidence of Vital Signs Monitoring with Wireless devices.

Although different systematic reviews show that the evidence to support the use of continuous measurement of vital signs with wireless devices is inconclusive (Cardona-Morrell et al., 2016; Leenen et al., 2020; Wells et al., 2022), there is a growing body of research supporting this claim. According to recent studies, this new technology offers an accurate record of vital signs (Jacobs et al., 2021; Posthuma, Visscher, et al., 2020; van der Stam, Mestrom, Scheerhoorn, et al., 2022) and support in the early detection of clinical adverse events (Breteler et al., 2018; Posthuma, Downey, et al., 2020). However, the sheer variety of technologies that are currently available, each with a specific set of functionalities, makes it challenging to address the reliability of these devices (Leenen et al., 2020). Another layer of difficulty is that many of the validation studies for these devices have been conducted with healthy volunteers and low-risk patients, making the results inapplicable to high-risk clinical settings (Breteler et al., 2018).

The majority of studies on reliability compare the precision of wearable sensors to other measurement methods like nursing measurements and bedside monitors (van Rossum et al., 2022; Wells et al., 2022). The results of these tests generally show that electronic devices can measure heart rate with good accuracy and with only small mean difference errors; however, the performance is rather subpar when measuring the respiratory rate (Wells et al., 2022). This unusual data could be the result of a bad wireless connection or the patient's level of movement, even though in-bed electronic devices or humans monitoring the RR exhibit different values.

To understand these devices better, it will be helpful to divide them into two categories. First off, there are more advanced devices with extensive sensors made for in-depth prehospital (ambulance) or clinical physiological monitoring, like the ViSi Mobile and WVSM Device. The monitoring capabilities of these devices are comparable to those of conventional intensive care units, enabling the gathering of extensive and detailed data. However, a different class of monitoring equipment prioritizes functionality for wireless ambulatory clinical monitoring. Since they only provide a small number of vital signs, these gadgets are appropriate for daily use. Chest patch sensors like SensiumVitals, VitalPatch, and Philips Biosensor are examples of such devices (Leenen et al., 2020).

The current study will use a wearable device called Healthdot, a chest patch developed by Phillips Healthcare to track the respiratory and heart rate of the patient. This specific device has been positively validated in clinical settings by two studies. Jacobs, et al. (2021) assessed the accuracy and precision of Healthdot in post-bariatric patients, comparing its measurements with standard electrocardiogram (ECG) and capnography measurements. The study found that Healthdot offers an accurate solution for both Heart Rate (HR) and Respiratory Rate (RR) measurements since 87.5% of the patients met the HR requirements and 92.3% met the RR requirements using a 5-minute average data. In the same way, Van der Stam, Mestrom, Scheerhoorn, et al., (2022) made the same comparison but this time with abdominal cancer patients finding better accurate results of more than 95% of precision in both vital signs. Since the presence of subcutaneous fat around the chest may cause accelerometer measurements to be inaccurate (Kant et al., 2022), the lower percentages presented Jacobs et al., (2021) study may be explained by the higher BMI of the bariatric study cohort.

b. *Analyzing wireless devices and clinical outcomes*

The performance of the wireless monitoring patch can be addressed by investigating their prediction of the early deterioration or improvement in clinical outcomes. In the study of Breteler et al. (2018) , the authors found evidence to support the use of wearable devices. Even though the study was carried out just with 31 patients of which just 11 developed an adverse event, the authors could find a significant number of abnormalities in every device tested before the actual adverse event. In specific, Atrial fibrillation (AF) was the most common event and it was recognizable due to a sudden increase in HR in all recordings by SensiumVitals and HealthPatch.

Other research with comparable artifacts discovered predictability in rehospitalization for heart failure (HF). The researchers (Stehlik et al., 2020) showed that a customized machine learning analytical platform supplied with patient monitoring data can accurately predict HF rehospitalization by using the wearable sensor (Vital Connect). The platform was able to identify the risk of hospitalization for HF worsening with 76.0% to 87.5% sensitivity and 85% specificity, depending on the method used to judge pre-event positive windows.

Finally, Van der Stam, Mestrom, Nienhuijs, et al.(2022) developed an early warning scoring system based on vital parameters measured by a Healthdot patch. A total of 103 were included in the study of which 29 experienced clinically relevant deterioration during the study period. The cutoff point for the remote early warning scoring (CREWS) was chosen to match the number of true positives obtained in current practice, which resulted in comparable sensitivity and specificity for the REWS compared with the counterpart of the conventional modified early warning system (MEWS).

In conclusion, there is a growing research body dedicated to analyzing the effectiveness, accuracy, and precision of the use of wearable monitoring devices. Many studies have shown that Healthdot patches and similar sensors perform well, establishing the feasibility of using this technology in clinical settings. Moreover, the potential improvement of using continuous monitoring to prevent the further development of adverse events is recognized. However, the current way to identify abnormalities and thresholds even with traditional early detection systems is still being adapted and developed.

c. *Limitation in the use of wireless devices for Early Warning Systems (EWS)*

Wireless devices hold great promise in healthcare for continuously monitoring patients' vital signs, allowing for the rapid identification and response to any alterations in comparison to intermittent measurements. While it would be ideal for every patient to receive intensive care unit (ICU)-style continuous monitoring, wearable devices offer a potential solution in this regard (Downey et al., 2017). Nevertheless, the application of EWS into this context still remains as a challenge due two main problems: to the amount of data loss in the records of this wireless devices, and the amount of false-alarms rates while applying the EWS with this remote continuous monitoring.

Regarding to the data quality problem of vital signs recorded by wireless devices, studies using remote devices have reported a data loss of 30 to 40% of the recordings (Jacobs et al., 2021). Although connectivity issues may contribute to this data loss, it remains unclear whether the devices fail to record information from specific populations. Similar to oximeters and other tools, the accuracy of measurements from these devices can be affected by factors such as skin color not being properly detected (Keller Matthew et al., 2022).

Hence, it is worth investigating whether this missing data could be related to other variables that hinder device performance. In a study assessing the accuracy of the Healthdot patch for continuously monitoring respiratory rate (RR) and heart rate (HR) in bariatric patients, the authors hypothesized that the required accuracy and precision using accelerometry in bariatric patients may be challenging due to their high Body-Mass-Index (BMI) and thicker layer of subcutaneous fat around the chest. While the measurements were found to be reliable, a significant proportion of low-quality data for heart rate was observed, around 34.5% (Jacobs et al., 2021). Although, this data loss could be attributed to errors in device placement or other malfunctions, the authors did not rule out the possibility that specific characteristics of the population studied were contributing factors.

On the other hand, regarding the number of false-negative alarms, this can be because EWS was developed within the context of intermittent vital sign measurements but not as a system for continuous monitoring. Differences in values between the two methods are expected, therefore there is a need for adjustments for the application of EWS in wireless continuous monitoring settings. For instance, the physical intermittent measurement can show different values due to what is referred as the "white coat syndrome" where vital signs can be elevated simply by the arousal effect of the nurse interaction (C. Downey et al., 2019).

Although existing literature has demonstrated inherent flaws in EWS alerts for certain populations, the challenge is amplified when measurements are continuous. Alarm fatigue, characterized by excessive false-positive alerts, has been identified as a key issue (Jeskey et al., 2011). To mitigate this challenge, numerous strategies have been developed, including adjusting thresholds based on population characteristics, night/daytime adjustments, or implementing adaptive thresholds based on pre-defined settings (e.g., pre-operation). While these approaches have shown improvements in alarm performance, none of the proposed solutions truly address the analysis of individual factors in the development of adaptive thresholds.

2.1.4 Association of individual factors (age, gender, BMI) and vital signs thresholds

Although most of the early warning systems in clinical deterioration use vital signs, the thresholds are developed solely on the information received from the intermittent measurements. This can present serious problems because what can be considered a normal vital sign deviation could differ by person depending on multiple factors (Ljunggren et al., 2016). In this section, evidence of individual factors that affect the vital signs trends will be discussed.

Vital signs tend to alter as a patient gets older. For instance, while blood pressure may rise as we age (Pinto, 2007), heart rate normally declines (Yeragani et al., 1997). In the same way, the respiratory rate can decrease with increasing age due to physiological changes to the respiratory system that affect the chest wall, the shape of the diaphragm, and the lung parenchyma itself (Chester & Rudolph, 2011).

Gender also has been shown to affect heart rates and respiratory rates. For instance, studies have shown that women tend to have lower heart rates compared to men due which can be due to differences in body size, physical activity, and hormone levels (Koenig & Thayer, 2016; Wei et al., 2017). Along the same line, there is research that supports that men tend to have a slightly higher respiratory rate compared to women (Lomauro & Aliverti, 2018).

In the same way, a difference in vital signs can be caused by several illnesses, including diabetes, heart disease, and respiratory problems. Individuals with chronic obstructive pulmonary disease (COPD) had higher respiratory rates compared to those without these conditions (Jensen et al., 2013). In the same way, heart rate can differ depending on the pre-health conditions. For instance, diabetes can cause a slower heart rate (Bassi et al., 2018).

All of the beforementioned factors can have an impact on the way that vital signs are used fot EWS, especially when they are the main input for deteriorative detection. Although there is awareness of the lack of sensitivity of general thresholds as decision support (Downey et al., 2017; Langsted et al., 2020), there are few attempts to introduce the individual variability inside ADS like the EWS. For instance, Van der Stam, Mestrom, Nienhuijs, et al., (2022) support investigating the role of specific individual factors like age, gender, and co-morbidities due to the high amount of false negative with the traditional EWS. Along the same lines, Langsted et al. (2020) found that Including Charlson Comorbidity Index status in EWS or adjusting for Charlson Comorbidity Index -status could increase the predictive value of the EWS in predicting 7-day mortality.

To conclude, it is important to incorporate individual factors in the analysis of vital signs. Due to the development of recent technology, there are technological opportunities to address the lack of incorporation individual factors in the analysis, namely by using the continuous monitoring of vital signs proportionate with wireless wearables devices. Such a device introduces a new opportunity for analysis and has the potential to increase decision-support systems like the EWS.

3. Conceptual framework

In this section, a conceptual model is provided of the key stages and components involved in a comprehensive Algorithmic Decision-Support system for public service delivery. The novelty of this model relies on tracking all the flow starting from the input through the transformation and the results of the algorithmic support systems. This approach highlights the importance of analyzing missing data and its interaction with other patient variables.

The primary objective of this analysis is to examine the data and outcomes of an Algorithm Decision-Support system, specifically the Early Warning System (EWS). To achieve this the study focuses on the measurement of vital signs using a wireless device and explores their implications within the EWS framework (Graph 1).

The model is divided into two stages. The first stage will show the retrieval of information from the wireless sensor and transformed into raw vital signs. In this stage, it is theorized that factors linked to the subject are affecting the amount of data retrieved from the wireless device. For explanatory purposes of this thesis, the factors will be divided into two types: contextual and individual. The first type of factor, contextual, relates to variables that are associated with the patient and her surroundings, such as the patient's posture when lying in bed and the degree of physical activity. In contrast, the second type of factors, individual, pertain to biological attributes such as gender and age, as well as relevant health indicators like the Body Mass Index.

As discussed in the literature review, the amount of data loss can be due different factors. While it is reasonable to think that activity level can lead to sweating and affect the adherence of the patch (Bent et al., 2020), on the other hand it is also possible that some positions can create additional noise, as certain body positions can create pressure where the patch is attached to the patient. Therefore, two hypotheses are formulated:

- H1: Contextual factors are associated with the loss of data in heart rate measured by the wireless device.
- H2: Contextual factors are associated with the loss of data in respiratory rate measured by the wireless device.

Similarly, the consideration of individual factors should not be disregarded when attempting to determine the cause of data loss. In this scenario, it is plausible that there may exist a potential correlation between gender, age, and body mass index with regards to the patterns observed in the data that is currently missing. The current body of literature has not yet provided sufficient evidence to support this claim. However, it is acknowledged as a potential hypothesis that specific populations may be associated with data loss (Jacobs et al., 2021).

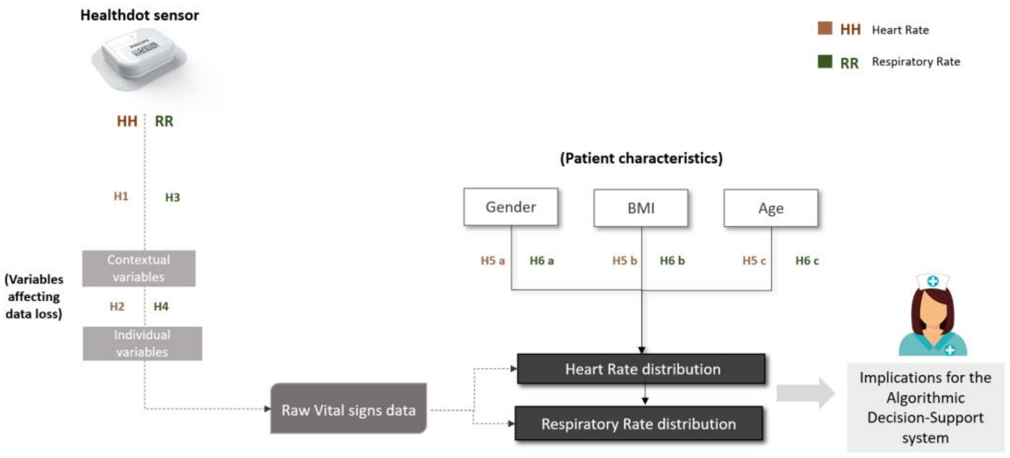
- H3: Individual factors are associated with the loss of data in heart rate measured by the wireless device.
- H4: Individual factors are associated with the loss of data in respiratory rate measured by the wireless device.

The second stage hypothesized that vital signs baselines are different for patients depending on individual characteristics such sex (Koenig & Thayer, 2016; Wei et al., 2017), age (Chester & Rudolph, 2011; Shindo et al., 2021; Zhang, 2007), and BMI (Littleton, 2012). The examination of these variables' relationship gain insights into how patient-specific variables may impact the interpretation of vital signs within the EWS. As discusses in the literature, individual specific characteristics gives different baselines and therefore requires specific thresholds. Therefore, two hypotheses are formulated:

- H5: Heart Rate presents different distributions depending on individual factors.
    - H5a Heart Rate presents different distributions depending on gender.
    - H5b Heart Rate presents different distributions depending on BMI.
    - H5c Heart Rate presents different distributions depending on age.
- H6: Respiratory Rate presents different distributions depending on individual factors.
    - H6a Respiratory Rate presents different distributions depending on gender.
    - H6b Respiratory Rate presents different distributions depending on BMI.
    - H6c Respiratory Rate presents different distributions depending on age.

Lastly, the model shows that any factor explained before and the process behind the ADS will have serious implications for the Decision-Making Support System, in this case, the EWS. Understanding the process and limitations is an essential step for optimizing the EWS and ensuring that it accounts for the diverse and unique aspects of each individual.

Graph 1: Conceptual Framework



4. Methodology

4.1 Data collection

The data for this study was gathered from Isala Hospital-Zwolle during the period from November 2021 to August 2022. Isala Hospital is a large regional hospital, encompassing multiple locations in the region across the Netherlands, with its primary facility situated in Zwolle. As part of the hospital's research projects on wearable devices, the information for this study was available and provided at the beginning of the research.

The study involved 403 patients who were subjected to continuous monitoring of their vital signs during their stay at the designated medical unit. The monitoring period's duration varied between 3.5 hours and 737 hours. In order to maintain uniformity, the application of the wearable device was restricted to individuals who had been hospitalized for a duration of at least 48 hours.

The wireless device used for retrieved the vital signs of the patients was the Heathdot patch manufactured by Philips Electronics Nederland B.V. This device is a small patch, 5x3 cm in size, is attached to the patient's lower left rib on the mid-clavicular line (Appendix 1) (Van der Stam, Mestrom, Scheerhoorn, et al., 2022). The Healthdot possess an accelerometer-based technology that process the motion signal to calculate heart rate (HR), Respiratory Rate (RR) and other measurements. The aforementioned data is stored within the system and subsequently transmitted at regular intervals of 5 minutes to a cloud server in the form of an aggregated average (Jacobs et al., 2021).

In this study, I used the information retrieved from the cloud server of the Heathdot. HR and RR are displayed in beats per minutes (bpm) and breaths per minutes(rpm) respectively. Additionally, physical activity is measured by the accelerometer and gives an activity level scale from 0 to 10. Furthermore, the device assesses the body position of the patient, providing a total of eight options for measurement. These options include lying on the abdomen, lying on the right side, lying on the left side, lying in a reversed position, being in an upright position, being in a backward position, being in a supine position, and being in a forward position (Appendix 2).

4.2 Data cleaning and pre-processing

This study initially included a dataset consisting of 403 patients and 416 series of vital signs to be analyzed. To ensure the consistency and reliability of the data, patients who passed away during the study period were subsequently removed from the dataset. This exclusion aims to eliminate additional variability that could potentially affect the study outcomes. As a result, the final dataset comprised of 392 series of vital signs collected from 381 patients. This exclusionary measure, which encompassed 5% of the population (equivalent to 22 patients), was made due to substantial dissimilarity in their biomarkers and the insufficient sample size to appropriately incorporate them as a control variable. As a result, the final dataset consists of 392 series of vital signs, collected from a cohort of 381 patients.

Demographics of the included patients are presented in Table 1, providing a comprehensive overview of the study population. However, out of the total amount of series with vital signs, 39 series did not contain recorded demographic information. As a result, the final dataset consists of 353 series with available individual variables.

Table 1 : Fundamental demographic information[2]

Gender	Patients	Obs	(%)
Female	146	151	42.80%
Male	198	202	57.20%
Total	344	353	

Age	Patients	Obs	Mean	Min	Max	Stand. Dev.
18-30	4	10,381	22	21	27	2
30-50	23	30,403	43	33	50	5
50-70	137	214,509	62	51	70	5
>70	176	242,955	79	71	96	6
Total	340	498,248	68	21	96	14

BMI	Patients	Obs	Mean	Min	Max	Stand. Dev.
BMI<25	117	191,964	22	14	25	2
25<BMI< 30	130	188,171	27	25	30	1
BMI>30	133	96,156	35	30	65	6
Total	380	476,291	26	14	65	6

[2] 39 series did not count with demographics values.

Department	Patients	Obs	(%)
Internal Medicine	89	93	26.30%
Lung disease	33	33	9.30%
Surgery	222	227	64.30%
Total	344	353	

4.3 Data analysis design

The study follows a quantitative design based on the recording of the vital signs of patients over a range of two days up to two weeks. This data will be used for correlational analysis, mean differences, distribution comparison, and regressions to assess the relationship between the data quality and the relationship between vital signs and patients' characteristics.

The study will follow two steps. The first step is to assess the quality of the data and find any link between the missing values, understood as an empty record or corrupted data, and any specific patient characteristic. To assess the missing patterns in the vital signs, the first step involves doing Little's chi-squared test to determine if the missing data is completely random (MCAR). If the likelihood of an observation being missing is equal across all cases, it is referred to be MCAR (Li, 2013).

Little's chi-squared test was introduced in 1998 as a multivariate test to examine the Missing Completely at Random (MCAR) assumption. This test assesses mean differences across subgroups that exhibit the same missing data pattern. It achieves this by comparing the observed variable means for each missing data pattern with the expected population means, which are estimated using the maximum likelihood estimates (Craig, 2010; Little, 1988) .

In the case of rejecting the null hypothesis MCAR, it is inferred that other variables are affecting into the missing patterns. To see if any observable variable is associated with the missing value, a logistic regression will be performed which will include as a dependent variable a dummy variable indicating 1 if the value in the vital sign is missing and 0 if it is not missing. For the independent variables, posture, activity level, gender, BMI and age will be included.

The first phase is critical because it determines whether there may be a systematic underrepresentation in the recording of wearable devices, which may discriminate against specific groups due to their inability to reliably capture vital signs. Because of this, the first step is to identify any potential bias in the total number of vital signs recorded. This is important to maintain internal validity at the moment of claiming a relationship between vital signs and patients' characteristics.

In the second phase, after the quality is assessed, the vital signs and their relationship with individual characteristics will be evaluated. For this, every patient characteristic will be divided into groups and a distribution comparison will be performed against each group. In this way, a t-test and ANOVA test will be used to assess how heart rate and respiratory rate data distributions are different depending on the most essential patient characteristics (gender, age, and body mass index).

With more than one million observations, it is assumed that the underlying distribution per group will behave as a normal function; however, it is expected that the mean and the standard deviation will differ in each group. The analysis focuses mostly on the differences in means, which following the central limit theory will behave as normally distributed if the sample size exceeds 30 observations per group (Table 1).

4.3 Analysis of relevant results

Regarding missing data, studies have demonstrated that a 10% bias is sufficient to significantly skew data results (Bennett, 2001; Lee & Huber, 2021). According to statistical guidance articles, if more than 40% of the data for crucial variables is missing, significant bias will be found in the analyses, and the results should only be viewed as hypothesis-generating (Dong & Peng, 2013; Jakobsen et al., 2017). As a result, for the purposes of this study, a moderate quantity of missing data is defined as between 10% and 40%, while a large quantity of missing data is defined as greater than 40%.

Regarding the clinically relevant results, the medical standard for identifying relevant differences in heart rate and respiratory rate is 10±10 beats per minute and 3±3 breaths per minute, respectively (Leenen et al., 2020). For the purpose of this study, it will be considered clinically relevant if the statistical significance is s (p 0.001) and if the difference agrees with the specified limits.

5. Results

In the subsequent chapter, the outcomes of the statistical analysis will be presented. This chapter will be divided into two distinct segments. The initial segment will assess the first four hypotheses concerning the quality of vital signs. Conversely, the second segment will explore the remaining hypotheses, focusing on the relationship between individual characteristics and vital sign distributions.

5.1 Quality of vital signs and missing values

5.1.1. Descriptive statistics

As stated before, the quality of data in the study is understood as the completeness of the data. In this regard a value is considered missing when there is a loss in the data set of vital signs and the records are not shown or are corrupted. From the total recordings of the dataset, heart rate represents 27% of the data loss, while respiratory rate showed data loss of 15%. In this regard, records of vital signs showed to be 16% more complete within respiratory rate in comparison of heart rate records.

Table 1: Total recording of Heart Rate and Respiratory rate

Recordings	HR		RR	
	n	%	n	%
Complete	404,419	73%	469,769	85%
Missing	148,165	27%	82,815	15%
Total	552 584	100%	552 584	100%

Table 2: Descriptive summary of Heart Rate and Respiratory Rate

Variable	Obs	Mean	Std. Dev.	Min	Max
Heart Rate	404419.00	77.78	14.25	31.00	190.00
Respiratory Rate	469769.00	18.33	4.80	5.00	54.00

To assess if this missing data has a special pattern related to observable variables, a Complete Missing at Random multivariate test was performed. The variables tested were Heart Rate and Respiratory Rate against activity level, position, BMI, age, and gender. The chi squared test showed a p-value less than 0.001 which is significant and lets us reject the null hypothesis of random missing (Table3). The rejection of hypothesis lets us infer that there is a possible association between the variables and the missing data.

Table 3: Complete Missing at Random testS

Missing data patterns heart rate position, activity, BMI, Age, and Gender	
Number of obs	552 583
Chi-square distance	75 205.8
Degrees of freedom	41
Prob > chi-square	<.001
Missing data patterns respiratory rate position, activity, BMI, Age, and Gender	
Number of obs	552 583
Chi-square distance	81 808.2
Degrees of freedom	39
Prob > chi-square	<.001

As discussed in the framework, two types of variables affecting data loss will be evaluated. The first group includes the contextual variables such as activity and position which are measured by the same wireless devices.

In Table 4, certain activity levels present a larger portion of missing vital sign data. When the activity level reached more than seven in the scale, around 90% of the heart and respiratory records are missing. The same effect is saw when the activity level is 0, were the amount of missing in heart rate and respiratory rate is 99% and 95% respectively. It is essential to note that the number of recordings for activity levels nine and ten, 222 and 66 respectively, are the shortest, which could reduce statistical power. The significance of these findings will be evaluated in the section that follows.

Table 4: Activity level and vital signs missing values.

Activity level	Total Recordings	Number of missings HR	Number of missings RR	% HR missings	% RR missings
0	10,431	10,372	9,862	99%	95%
1	4,177	1,794	1,143	43%	27%
2	32,571	3,395	445	10%	1%
3	154,251	17,643	859	11%	1%
4	181,171	39,657	9,868	22%	5%
5	85,770	37,765	26,747	44%	31%
6	33,695	23,164	20,735	69%	62%
7	12,910	11,936	10,873	92%	84%
8	2,150	2,135	2,000	99%	93%
9	222	222	215	100%	97%
10	66	66	66	100%	100%
Total	517,414	148,149	82,813	29%	16%

Regarding to the posture as a contextual variable, it is seen that when the position are "lying on belly", "Upright" and "Forward" there is a large amount of missing data (Table 5). While in heart rate, the number of missing recordings obtained during lying in belly reached the 62% this number is lower for heart rate being 49%. During the "Upright" position, the level of missing reached 60% for heart rate and 52% for respiratory rate. Finally, during the forward position, the level of missing values for heart rate was 58% and the 49% for respiratory rate. This lead to inquire a possible relationship between the position and the missing data in vital signs, however the significancy of these results will be evaluated in the next section.

Table 5: Position and vital signs missing values.

Position	Total Recordings (ascendant)	Number of missings HR	Number of missings RR	% HR missings	% RR missings
Supine position	221,418	49,760	18,820	22%	8%
Backward	131,553	39,422	19,241	30%	15%
Upright	55,342	33,314	28,982	60%	52%
Lying right	49,094	8,078	6,605	16%	13%
Lying left	43,237	7,784	1,454	18%	3%
Lying on belly	8,707	5,431	4,282	62%	49%
Forward	4,887	2,821	2,406	58%	49%
Other way around	3,176	1,539	1,023	48%	32%
Total	517,414	148,149	82,813	29%	16%

In relation to the distribution of missing values across individual variables, patterns of moderate percentages of missing data can be seen through the groups of BMI and age. Specifically, in the case of BMI, the percentage of missing goes from 18% (<25) for the lower level to 35% to the largest level (>30). A similar pattern is observed in the distribution of missing data by age, with the youngest individuals exhibiting the lowest proportion of missing data at 8%, while the oldest individuals had a higher proportion of missing data at 32%. In contrast, the respiratory rate exhibited minimal variation, remaining relatively constant.

Table 6: Individual variables and vital signs missing values.

Variables	Total of Recordings	Number of missing HR	Number of missing RR	% HR missings	% RR missings
Gender					
Female	207,986	51,124	32,708	25%	16%
Male	298,430	82,505	44,162	28%	15%
Total	506,416	133,629	76,870	26%	15%
BMI					
BMI less than 25	191,964	34,886	29,168	18%	15%
BMI more than 25 but less than 30	188,171	53,716	28,095	28%	15%
BMI more than 30	172,449	60,547	25,550	35%	15%
Total	552,584	149,149	82,813	27%	15%
Age					
0-30 years	10,381	831	704	8%	7%
30-50 years	30,403	4,891	3,050	16%	10%
50-70 years	214,509	47,945	27,704	22%	13%
More than 70 years	242,955	77,546	42,845	32%	18%
Total	498,248	143,307	74,303	26%	15%

The observation implies that there may exist a potential association between individual variables and the occurrence of missing data in the context of heart rate. However, it is noteworthy that such a relationship is not as readily apparent when considering respiratory rate. The forthcoming section will entail the execution of a logistic regression analysis to evaluate the statistical significance and effect of the variables evaluated.

5.1.2. Association between the contextual and individual variables with missing values.

Heart Rate

To examine the hypotheses ranging from 1 to 4, a logistic regression analysis was conducted. For this regression analysis, some variables were aggregated to make the inference easier to understand (Table 7). In this way, BMI was grouped into three categories following medical standards (Nuttall, 2015). Additionally, the age was organized into four groups: less than 30, 30 to 50, 50 to 70, and more than 70 years old.

The regression revealed a significant association between higher activity levels and the presence of missing values in Heart Rate measurements, as indicated by a predictive margin of more than 0.91 for level 7 and 0.99 for level 8, both with a z-value lower than 0.001 (Table 7). Similarly, specific body positions, namely "Lying on Belly," "Upright," and "Forward," demonstrated predictive margins of 0.47, 0.32, and 0.30, respectively, with z-values significantly below 0.001. These findings suggest that the occurrence of missing data in heart rate is positively associated with certain positions and higher levels of activity. Consequently, the results provide support for **Hypothesis 1** that states that contextual factors are associated with the loss of data in heart rate measured by the wireless device.

Regarding the individual variables, the logistic regression analysis also included the predictive margins for gender, age, and BMI (Table 7). The results indicate that the male gender exhibited a predictive margin of 0.27, which is 0.03 points higher than that of the female gender. This suggests that both genders have similar effects on the probability of resulting in heart rate missing values and there is not a relevant effect of gender on missing heart rate.

In terms of BMI, the group with a BMI greater than 30 demonstrated a predictive margin of 0.41, which is 0.20 points larger than the group with a normal BMI, and 0.11 points larger than the overweight BMI group. These findings suggest that individuals with higher BMI values have a greater likelihood of experiencing missing values in heart rate data compared to those with lower BMI values.

Furthermore, when examining the age groups, the predictive margin was found to be higher for the "more than 70 years" age group, with a value of 0.34. This is more than 0.20 points larger than the predictive margins observed for the youngest age group.

These results indicate that individuals in the older age group are more likely to have missing values in heart rate data compared to their younger counterparts.

The information presented about the individual variables indicates that the BMI and age increase the probability of missing values in heart rate. For BMI, the predicted probability is 41% when the patient has a BMI level greater than 30, and for age the predicted probability is 34% when the patients is older than 70 years. These results support **Hypothesis 2** that states individual factors are associated with the loss of data in heart rate measured by the wireless device.

Respiratory Rate

Regarding to the Respiratory Rate, the regression analysis revealed a significant association between higher activity levels and the presence of missing values, as indicated by a predictive margin of more than 0.7 for levels 7, 8 and 9 and a z-value lower than 0.001 (Table 7). On the contrary, the lowest activity level recorded, the level zero, showed a 0.93 of predictive margin for the heart rate. On the regards of specific body positions, it is seen that “Lying on Belly” had the higher predictive margin in the group with a value of 0.24 and a z-values significantly below 0.001, and it is 0.15 points superior that “lying on left” (the category with less predictive margin). These findings suggest that the occurrence of missing data in Respiratory Rate is positively associated with certain levels of activity and positions. Consequently, the results provide support for **Hypothesis 3** that states that contextual factors are associated with the loss of data in respiratory rate measured by the wireless device.

Regarding the individual variables, the results indicate that the female gender exhibited a predictive margin of 0.17 which is the 0.01 point bigger than the male gender, showing no relevance difference in the probability between both groups (Table 7). The same results are observed with the BMI variable, were the predictive margins for all groups reach around 0.16 showing a low effect of this variable on the probability of missing values in the respiratory rate. When examining the age group, none of the groups have a predictive value more than 0.17, showing here too a week relationship between the age and respiratory rate missing values.

The information presented about the individual variables indicates a low association between the individual variables, less than 17% of predictive probability, and the missing values in respiratory rate. In this regard, the evidence **reported does not support** the **Hypothesis 4** that states individual factors are associated with the loss of data in respiratory rate measured by the wireless device.

Table 7: Logit regression and predicted margin of missing values vital signs against position, activity and individual variables

Variables	Missing of Heart Rate		Missing Respiratory Rate	
	Margin	P>z	Margin	P>z
Position				
Lying on Belly	0.457	>0.001	0.23	>0.001
Lying right	0.207	>0.001	0.20	>0.001
Lying left	0.260	>0.001	0.09	>0.001
Other way around	0.264	>0.001	0.15	>0.001
Upright	0.305	>0.001	0.18	>0.001
Backward	0.312	>0.001	0.16	>0.001
Supine position	0.269	>0.001	0.14	>0.001
Forward	0.296	>0.001	0.16	>0.001
Activity level				
0	0.99	>0.001	0.93	>0.001
1	0.34	>0.001	0.29	>0.001
2	0.10	>0.001	0.02	>0.001
3	0.12	>0.001	0.01	>0.001
4	0.22	>0.001	0.06	>0.001
5	0.41	>0.001	0.30	>0.001
6	0.64	>0.001	0.56	>0.001
7	0.92	>0.001	0.79	>0.001
8	0.99	>0.001	0.91	>0.001
9			0.96	>0.001
10				
Gender				
Female	0.26	>0.001	0.16	>0.001
Male	0.30	>0.001	0.16	>0.001
BMI				
BMI less than 25	0.20	>0.001	0.17	>0.001
BMI more than 25 but less than 30	0.28	>0.001	0.15	>0.001
BMI more than 30	0.40	>0.001	0.15	>0.001
Age				
0-30 years	0.13	>0.001	0.08	>0.001
30-50 years	0.16	>0.001	0.13	>0.001
50-70 years	0.25	>0.001	0.14	>0.001
More than 70 years	0.33	>0.001	0.18	>0.001

To summarize, it is concluded that the quality of data is a problem specially for the Heart Rate vital signs, where around the 27% of the data is missing in comparison with the 15% in Respiratory Rate. In assessing if these missing values were related to any observable variable in the data set it was found that contextual variables, activity, and position, did have an association with the missing values in both vital signs (Hypothesis 1 and Hypothesis 3). In specific, "lying in belly" position and high activity had the biggest probability on increasing the missing values observed in both vital signs. On the other hand, individual variables just showed to be associated in the heart rate case, where BMI and Age were the most related to missing values (Hypothesis 2).

5.2 Individual characteristics and vital sign

a. Gender

To investigate any potential gender-based variations in heart rate, a T-test was conducted (Table 7). The findings demonstrated that there are significant differences between male and female heart rates ($p<0.001$). The female heart rate is higher, although the difference is just about one point, being 78.5 point for female and 77.3 for male. Despite being statistically significant, the difference of one point base on gender is not clinically relevant because the difference is lower than the pre-establish range of +10,-10 bpm. In consequence, there is insufficient evidence to support the claim that heart rate values distribute differently depending on gender (sub-hypothesis 5 a).

Regarding the respiratory rate, a t-test was performed to find differences between female and male groups. It was found that the female has a mean respiratory value of 18.1 while the male has a mean value of 18.5, showing a difference of -0.4. However, although the difference is statistically significant ($p<0.001$), the size of the different is not clinically relevant as the difference is lower than the pre-establish of +3,-3 rpm margin. As a result, I conclude that there is **insufficient evidence** to support the claim that respiratory rate values distribute differently depending on gender (sub-hypothesis 6a).

Table 8: Mean test of vital signs values by gender

	Heart Rate values					Respiratory Rate values				
Gender	**Mean**	**Std. Err.**	**[95% Conf. Interval]**		**Std. Dev.**	**Mean**	**Std. Err.**	**[95% Conf. Interval]**		**Std. Dev.**
Female	78.50	0.03	78.39	78.52	12.39	18.10	0.01	18.06	18.11	5.15
Male	77.30	0.03	77.21	77.34	15.16	18.50	0.01	18.51	18.55	4.66
Difference	1.2 <0.001					-0.4 <0.001				

Figure 1: Distribution of Heart Rate by Gender

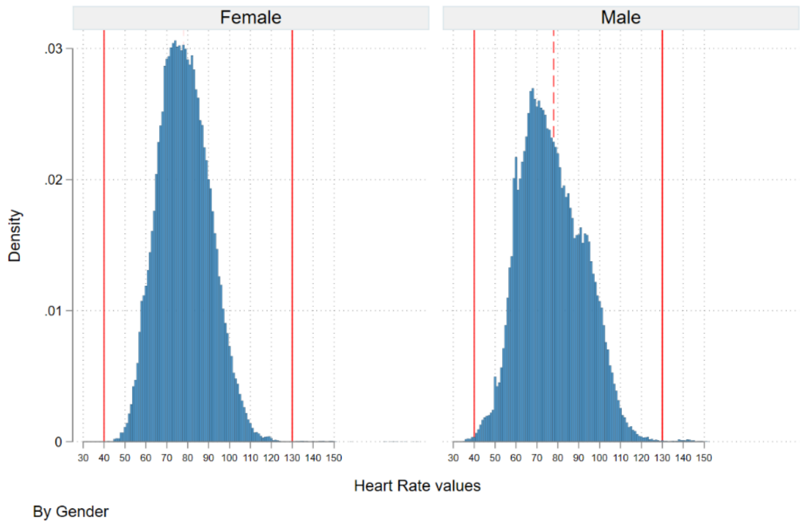
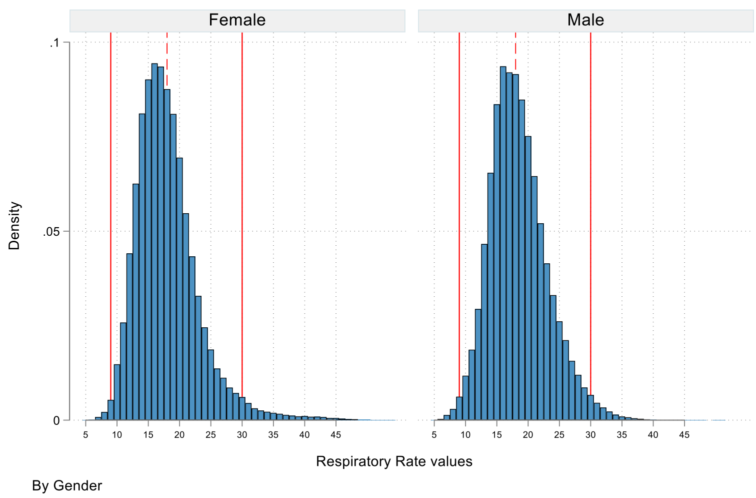


Figure 2: Distribution of Respiratory Rate by Gender



b. BMI

A one-way ANOVA was performed to examine the potential differences in heart rate among groups with different BMI classifications. The participants were categorized into three groups: BMI<25 (n = 157,075), 25<BMI<30 (n = 135,451), and BMI>30 (n = 111,893). The results indicated a statistically significant difference between the groups, as evidenced by the one-way ANOVA ($F = 2038.27$, $p < .0001$).

Further analysis using a Tukey post-hoc test revealed statistically significant differences in heart rate between certain pairs of groups (Table 8). Specifically, the heart rate was significantly higher in the 25<BMI<30 group compared to the BMI<25 group (mean difference = 3.06, $p < .0001$), indicating a contrast effect between these groups. However, this effect is not large enough to be consider clinically relevant considering the pre-established margin of +10, -10 (Leenen et al., 2020). In this way, there is not enough evidence that supports sub-hypothesis 5 b, which posits that heart rate exhibits distinct distributions based on BMI.

Regarding respiratory rate, the one-way ANOVA analysis revealed a statistically significant difference between the groups ($F = 3398.25$, $p < .0001$). Post-hoc analysis using the Tukey test indicated a statistically significant difference in respiratory rate between the BMI>30 group and the BMI<25 group (mean difference = 1.38, $p < .0001$), indicating a notable contrast effect between these two groups. Furthermore, there were statistically significant differences observed between the other groups as well, although the effect sizes were comparatively smaller.

To synthesize, the findings for respiratory rate show a different distribution per BMI classifications, however the effects were considered not clinically relevant. The normal group (BMI<25) showed a bigger rate compared with the overweighted group (25<BMI<30); however, the effect was considerable smaller, around 1.4 point. This effect is not considered relevant followed the pre-established standard of +-3 rpm (Leenen et al., 2020). The findings do not support the sub-hypothesis 6 b, which suggests that respiratory rate displays varied distributions patterns depending on BMI.

Table 9: Mean test for heart rate values by BMI

BMI	Mean	Sd
BMI <25	76	14.91
25<BMI< 30	79.06	12.80
BMI>30	78.72	14.72
Bartlett's test for equal variances	*(0.000)****	

Tukey's test	Contrast	Std. Err.	t	P>t	[95% Conf. Interval]	
25<BMI< 30 vs BMI<25	3.06	0.05	58.25	<0.001	2.94	3.19
BMI>30 vs BMI<25	2.72	0.06	49.05	<0.001	2.59	2.85
BMI>30 vs 25<BMI< 30	-0.34	0.06	-5.97	<0.001	-0.48	-0.21

Table 10: Mean test for respiratory rate values by BMI

BMI	Mean	sd
BMI <25	17.58	4.5520799
25<BMI< 30	18.5	4.8579999
BMI>30	18.96	4.8913531
Bartlett's test for equal variances	*(0.000)****	

Tukey's test	Contrast	Std. Err.	t	P>t	[95% Conf. Interval]	
25<BMI< 30 vs BMI<25	0.92	0.02	54.75	<0.001	0.88	0.96
BMI>30 vs BMI<25	1.38	0.02	80.45	<0.001	1.34	1.42
BMI>30 vs 25<BMI< 30	0.46	0.02	26.79	<0.001	0.42	0.50

Figure 3: Distribution of Heart Rate by BMI

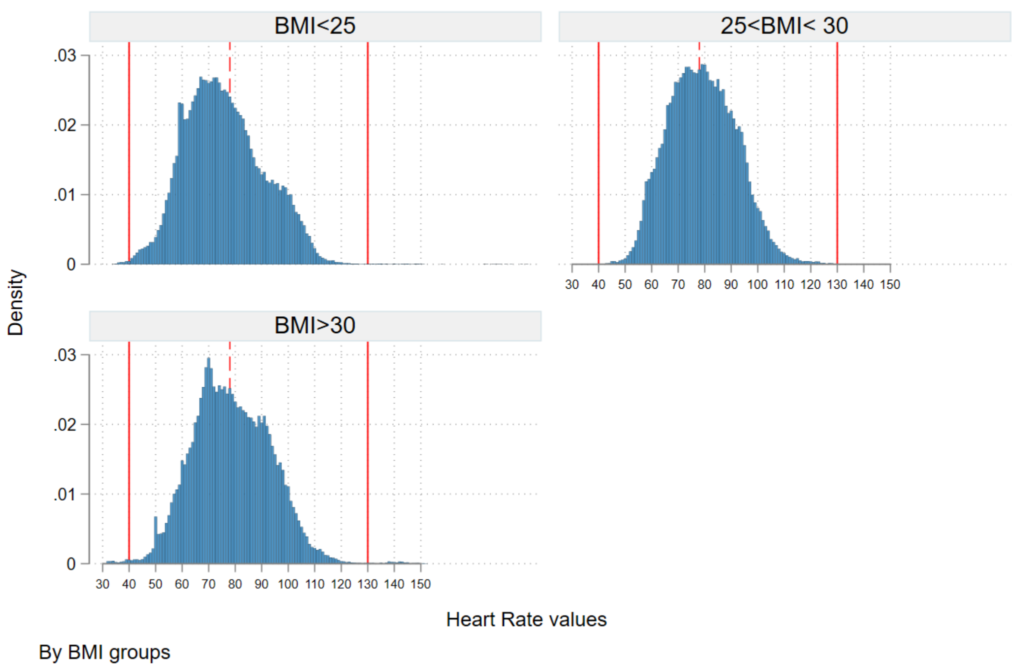
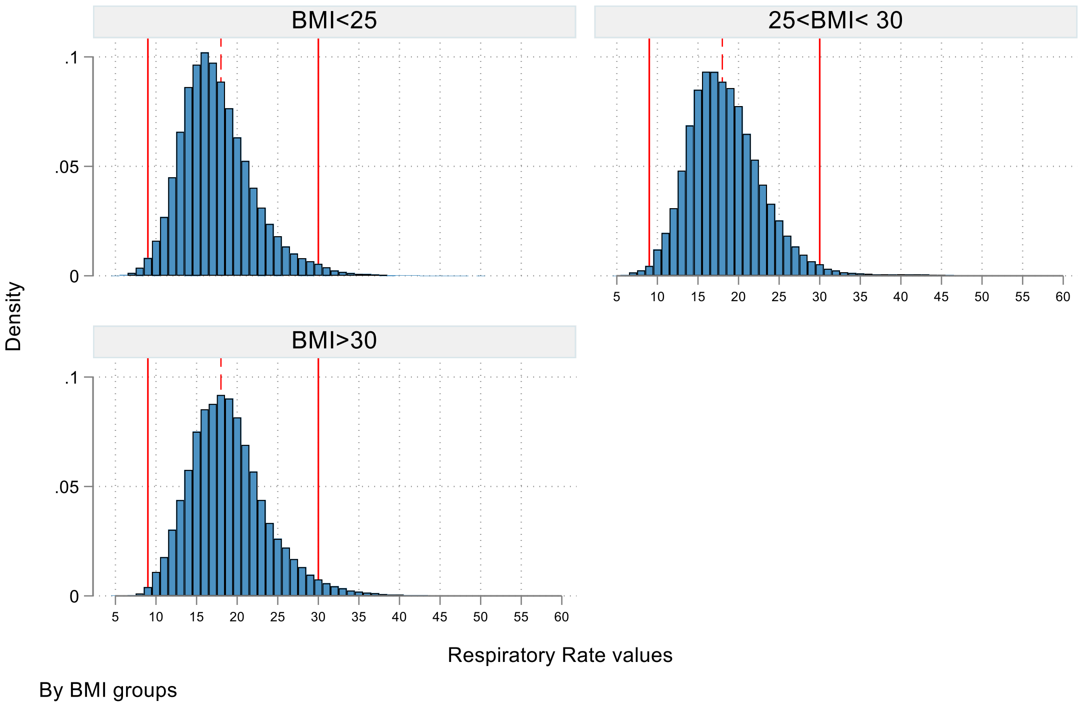


Figure 4: Distribution of Respiratory Rate by BMI



c. Age

A one-way ANOVA was performed to examine the potential differences in heart rate among groups with different age classifications. The participants were categorized into four groups: Age<30 , 30<Age<50 , and 50<Age<70, and Age>70. The results indicated a statistically significant difference between the groups, as evidenced by the one-way ANOVA ($F = 6537.58$, $p < .0001$).

The results showed that heart rate variates depending on different group ages (Table 11). Specifically, the bigger size and significant effects were found when the age groups were compared to the younger group " between 18 to 30 years old". In this regard, the younger group has a hear rate of 12 bpm bigger compared with the people in ages between 30 to 50 ($p < .0001$). A similar effect was found when the comparison of the same young group was between people with ages in the range 50–70 (difference of 18 bpm, $p < .0001$) and people who were more than 70 years old (difference of 19 bpm, $p < .0001$). Because these differences are bigger than 10 points in bpm, this result is considered statistically and clinically relevant(Leenen et al., 2020) . This finding supports the sub-hypothesis 5c that heart rate presents a different distribution depending on age.

The results of the ANOVA test revealed a statistically significant difference in respiratory rate among the groups ($F = 1659.27$, $p < 0.001$). However, the contrast effect sizes were relatively smaller(Table 12). Only one comparison, specifically between the "more than 70 years old" and "30 to 50 years old" age groups, showed a difference in respiratory rate exceeding one unit. Although there were significant differences in respiratory rate observed between age groups ($p < 0.001$), the magnitude of these differences was relatively small and did not pass through the pre-established margin of +3,-3 rpm (Leenen et al., 2020).

Therefore, the sub-hypothesis that states respiratory rate presents different d depending on age (sub hypothesis 6 c) cannot be supported based on the findings. Despite the statistical significance, the relevance of the observed differences in respiratory rate between age groups may be limited.

Table 11:Mean test for heart rate values by Age

AGE	Mean	sd
18-30 years	94.78	10.50
30-50 years	82.50	11.95
50-70 years	76.90	13.93
More than 70 years	76.61	13.86
Bartlett's test for equal variances	*(0.000)****	

Tukey's test	Contrast	Std. Err.	t	P>t	[95% Conf. Interval]	
30-50 VS <30	-12.27	0.16	-74.7	>0.001	-12.69	-11.85
50-70 VS <30	-17.86	0.14	-124.01	>0.001	-18.23	-17.49
> 70 VS <30	-18.17	0.14	-126.16	>0.001	-18.54	-17.80
50-70 VS 30-50	-5.59	0.09	-60.79	>0.001	-5.83	-5.36
> 70 VS 30-50	-5.91	0.09	-64.16	>0.001	-6.14	-5.67
> 70 VS 50-70	-0.31	0.05	-6.58	>0.001	-0.43	-0.19

Table 12: Mean test for respiratory values by age

BMI	Mean	sd
18-30 years	18.51	4.92
30-50 years	17.71	4.77
50-70 years	18.70	4.98
More than 70 years	18.84	4.61
Bartlett's test for equal variances	*(0.000)****	

Tukey's test	Contrast	Std. Err.	t	P>t	[95% Conf. Interval]	
30-50 VS <30	-0.79	0.05	-16.03	>0.001	-0.92	-0.67
50-70 VS <30	0.20	0.05	3.79	>0.001	0.06	0.33
> 70 VS <30	0.33	0.05	6.78	>0.001	0.21	0.46
50-70 VS 30-50	0.99	0.02	44.19	>0.001	0.93	1.05
> 70 VS 30-50	1.13	0.02	72.47	>0.001	1.09	1.17
> 70 VS 50-70	0.14	0.02	6.25	>0.001	0.08	0.19

Figure 5: Distribution of Heart Rate by Age

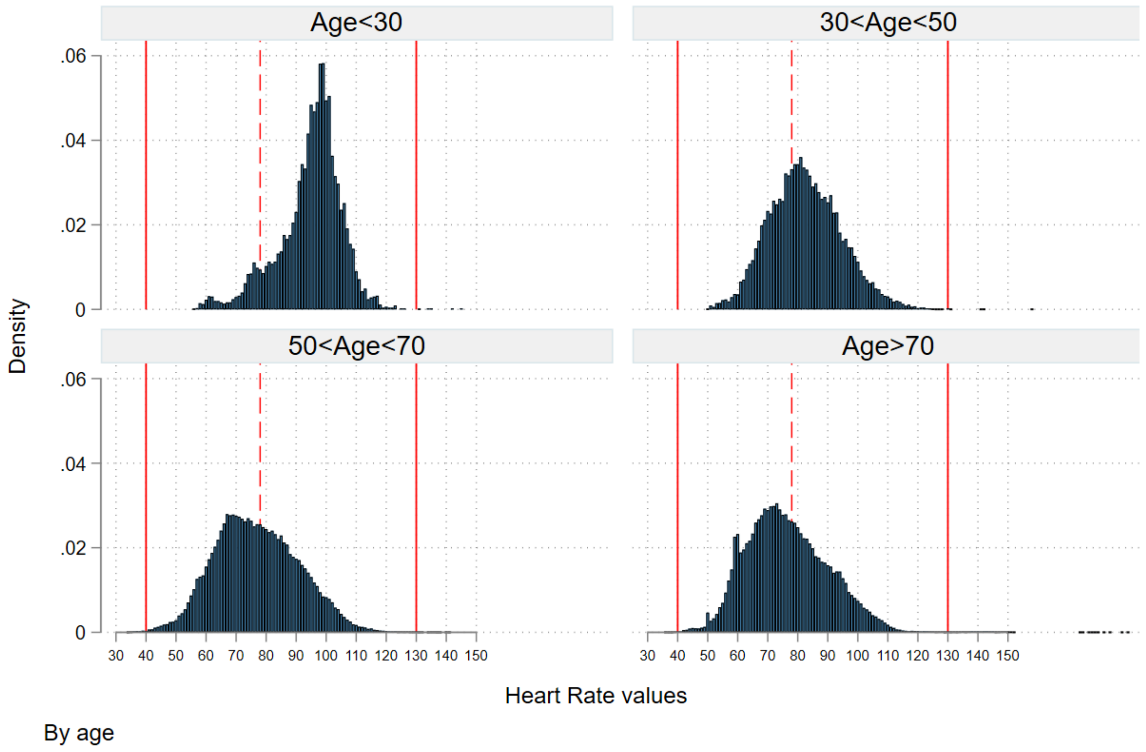
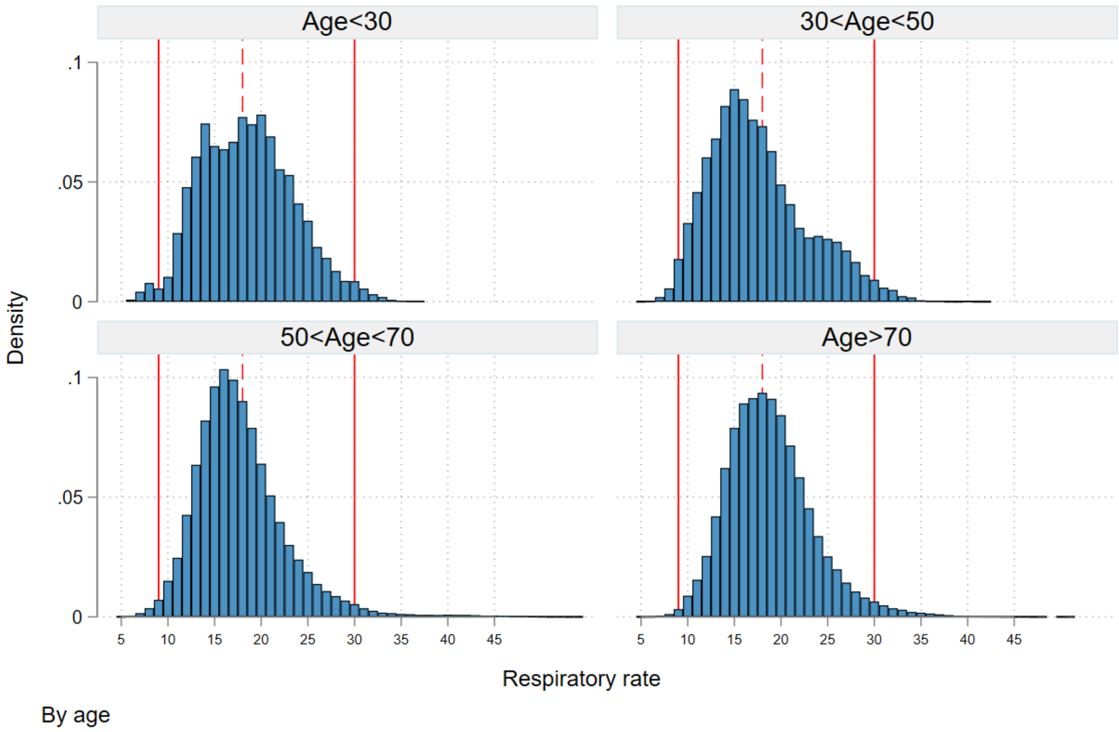


Figure 6: Distribution of Respiratory Rate by Age



In summary, the findings of these chapter suggest that heart rate distributions exhibit statistically and relevant significant differences depending on individual factor in the case of age groups. However, respiratory rate did not demonstrate significant differences based on individual factors. Consequently, sub-hypotheses 5a, 5b, 6a, 6b and 6c were rejected, while sub-hypothesis 5c was supported.

6. Discussion

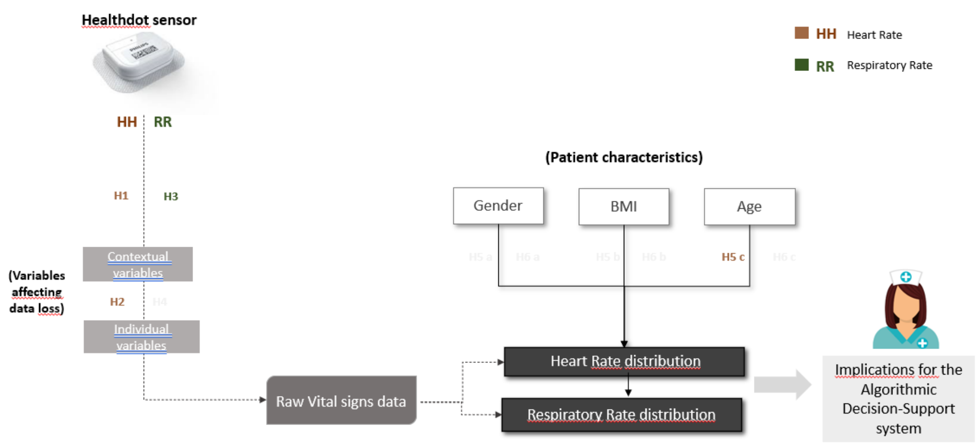
The primary aimed of this study was to investigate the utilization of an Algorithmic Decision-Support system for delivering public services, focusing on continuous vital sign monitoring within a medical context. Specifically, the study seeks to evaluate the data quality employed in the ADS system and understand to what extent do these systems effectively capture and represent individual characteristics. This study utilized the Healthdot Path, a wireless device capable of continuously monitoring vital signs and tracking both heart rate and respiratory rate in patients.

In this section, these results will be discussed and put in context to understand its implications. The interpretation of the results will follow the research questions , then the practical implication will be discussed in the light of Heath Sector and Public Administration. Finally, the limitations of the study and further research lines will be addressed.

6.1 Summary of the results

This research aimed to investigate the quality of data used for Algorithmic Decision-Support system and understand to what extent do these systems effectively capture and represent individual characteristics. To fulfill this objective, three research questions were formulated, each accompanied by corresponding hypotheses. Following the analysis, four of the hypotheses were upheld, substantiating the envisaged correlations between variables. A visual depiction of these outcomes is illustrated in Figure 7.

gure 7: Hypothesis supported in the study.



It was found that contextual variables that include certain positions and a high activity level have a strong association with the data loss for heart rate (H1, H2). For respiratory rate, it was found that higher activity levels were associated with data loss patterns (H3). Regarding the representation of individual characteristics, it was found that heart rate showed a different distribution when age was less than 30 compared with other age groups (H5c). Respiratory rate showed no difference in distribution depending on individual characteristics.

These results bear substantial implications for the utilization of wearable devices and decision support systems, such as Early Warning Systems (EWS), within the healthcare realm. The study underscores that EWS, often reliant on generic threshold values, might inadequately accommodate individual deviations in vital signs. Consequently, a one-size-fits-all approach may not be optimal, and a more tailored and personalized approach to EWS could improve patient outcomes. A more comprehensive exploration of these implications will be presented in the subsequent section.

6.2 Interpretation of Results

By following the logical order of the research questions, this chapter aims to provide a thorough and perceptive explanation of the obtained results, ultimately fostering a deeper comprehension of the implications and significance of the study's findings.

(1) *What is the quality of vital signals recorded by wearable devices, and is there a correlation between data loss and patient-related factors?*

Although some prior research on the Heathdot device has suggested connectivity issues or malfunctions as potential explanations for missing data, no definitive conclusions have been reached through empirical examination of this hypothesis (Jacobs et al., 2021). Thus, this thesis aimed to examine whether these missing patterns were entirely random or if they exhibited associations with observable variables.

Regarding the hypotheses H1 and H3, my results indicate a significant association between contextual variables, such as activity level and position, and the occurrence of data loss in both vital signs, respiratory rate and heart rate. Specifically, the results from our study revealed that activity levels beyond 7 were strongly correlated with a substantial increase in the likelihood of missing data for both respiratory rate (over 80%) and heart rate (91%) (Table 7).

However, interpreting these outcomes in practical terms poses some challenges. As outlined in the guidelines provided by Phillips (Philips Electronics Nederland B.V, 2020), an activity level of 0 signifies a state of rest, whereas an activity level of 10 corresponds to "maximum activity” . Unfortunately, the lack of additional references or contextual information hinders a comprehensive understanding of this activity scale's interpretation. Hence, there is still ambiguity regarding the extent to which "maximum activity" encompasses routine tasks such as walking to the bathroom or engaging in conversation. It is crucial to note that the patients in our study were hospitalized, with most of them requiring post-surgical intervention, implying that extreme physical exertion or strenuous activities were not plausible scenarios.

More research should be done to understand what this activity level scale means. Especially in situations where the devices are used for monitoring patients outside the hospital. The missing data at certain activity levels can result in not necessarily having a continuous monitoring of the patient and losing the option of getting an alarm when the patient requires attention. The loss of data can pose a critical problem, and a complete understanding of it by the medical stuff can help them to take better decisions. In recent years, research on motor artifacts in wireless vital signs devices has increased (An et al., 2022; Bent et al., 2020; Shcherbina et al., 2017). Although it involves different types of devices like exercise wrist bands and some medical devices, it is becoming evident that there are problems related to high activity levels and the potential of these devices to keep their performance in those instances.

Another interesting point that is also observed in the results, is that when activity was zero, the likelihood of missing data was high (99% for heart rate and 93% for respiratory rate). This finding appears to contradict the initial assumption that vital sign data will perform better at low activity levels. However, upon consulting with the medical team, it became apparent that the system may assign a zero value when the patch is not entirely and securely attached. Thus, this outcome could be indicative of a malfunction in the patch and the subsequent data recording process. Despite the absence of a definitive explanation, this observation warrants consideration and may have implications for the overall reliability and accuracy of the system's performance.

With regard to the hypotheses H2 and H4, our findings indicate a significant association between individual variables and the occurrence of data loss in heart rate (H2), but no significant association was found for respiratory rate (H4). On the side of Heart rate, upon certain level of BMI the amount the probability of data loss increased reaching to 40% when the BMI was more than 30 (Obesity). In comparison with the other two groups BMI<25 (Normal) and 25<BMI<30 (Overweight), the probability of getting data loss under obesity increased in 100% and in 40%respectively (Table 7). These results can serve as an indication of technical problems with this wireless device in recording certain populations.

Previous studies have assessed the accuracy of Healthdot measurements in bariatric patients, typically encompassing individuals with a BMI of 30 or higher. While these investigations did not reveal any statistically significant disparities between the recordings obtained via the Healthdot device and the hospital's designated gold standard, a noteworthy concern surfaced concerning the prevalence of missing data in heart rate recordings, amounting to approximately 35% of the total records (Jacobs et al., 2021).

This issue assumes critical importance as it brings to light a notable limitation of the aforementioned study: its exclusive focus on evaluating the accuracy of the available data, without accounting for the role of data loss and potential biases it may introduce in assessing the devices' accuracy. Addressing the issue of missing data is of utmost significance, as such omissions can potentially distort the overall evaluation of the device's precision and lead to skewed conclusions.

Finally, as an answer to the research question, the quality of data understood as data completeness is not the same for all the recordings. It depends on contextual variables such activity and position, as well as with individual variables such BMI. It is seen that the quality of data is worst when the activity levels are high and when patients present BMI levels more than 30.

(2) How do the heart rate and respiratory rate distributions differ between the most important patient characteristics, such as gender, age and body mass index?

The aim of this question was to analyze if the distribution of vital signs was similar across the different individual characteristics or if there were important differences in the distribution that led us to rethink the use of general standards in assessing the vital signs.

In the examination of Heart Rate data, the results revealed a marginal but significant difference of approximately 1.2 bpm between male and female individuals; however, this disparity did not yield clinical relevance (H5a). It is important to acknowledge that, due to the large volume of data in this study, even the smallest differences in distributions may register as statistically significant. However, when compared to the clinically relevant pre-established threshold of +10, -10 bpm (Leenen et al., 2020), these differences may not carry practical relevance. Likewise, when analyzing heart rate among different groups of Body Mass Index (H5b), non-substantive differences were observed, which may be attributed to the machine's inherent recording error rather than genuine variations in distribution.

In the analysis of continuous vital sign distributions, the only instance where statistically significant and practically relevant differences were observed pertained to the comparison between different age groups. Specifically, younger age groups exhibited approximately 18 bpm higher heart rate values than older age groups (H5c). According to existing literature, there is evidence to suggest that heart rate exhibits a decline as individuals progress in age (Yeragani et al., 1997). The existing guidelines delineate varying ranges of normal vital signs for pediatric populations (Fleming et al., 2011). However, in the context of adult patients, it is worth noting that there are currently no specific guidelines that propose distinct acceptable ranges based on age groups other than the broader range of 60 to 100 bpm (Avram et al., 2019). This observation is particularly significant considering that normal aging has been consistently linked to a gradual decline in cardiac vagal modulation, which serves as a plausible explanation for the observed variations (Bonnemeier et al., 2003).

On the side of respiratory rate, no individual characteristics such as gender, BMI and age had relevant results. In the aspect of gender (H6a), although the literature tends to indicate a gender difference in the RR, often sex differences can, in fact, be attributed to scale, as women are generally smaller than men (Lomauro & Aliverti, 2018). It is plausible that the absence of gender effects observed in this study could be attributable to the comparatively homogeneous population, in which there may not be significant differences in body size. To test this hypothesis regarding the potential confounding effect of body size on gender differences in RR, height would have been required as a control variable. Unfortunately, this study lacked data on participants' height, which represents a significant limitation.

In regard to BMI (H6b), the literature supports that with higher levels of BMI, the respiratory rate increases (Littleton, 2012); however, this effect is not seen in this thesis. One reason why this result was obtained is because the population in the research did not have enough representation for all levels of BMI, especially for the ones over 40 points, which were just 16% of the dataset. Studies stating differences in the respiratory rate patterns found differences of around 3 to 10 rpm compared with the normal subjects when the population in the studies presented morbid obesity (BMI > 40). For this reason, it can be hypothesized that with the current population used for this thesis, there were no significant differences, but it is necessary to conduct research on a more diverse population that includes representation of all levels of BMI.

Finally, with regard to age (H6c), the results on respiratory rate showed that there were no specific effect depending on respiratory rate. This finding contrasts with existing literature, which has generally supported distinct RR values between young adults (aged 18 to 70 years) and older adults (over 70 years) (Takayama et al., 2019). However, with this specific population this effect has not been detected. Small changes in respiratory rate are important however, the precision of the Healthdot measuring these vital signs has been putting into questions (Wells et al., 2022).

Finally, it is important to notice that this thesis just found a difference in distributions for the age group in heart rate. These results are constrained by the type of population we had, the type of device we used, and the bias toward lower activity levels. This does not rule out the possibility that with a more representative and diverse group, as well as all the activity levels in the range, these results could have been different. More research should be done in this regard to analyze these individual characteristics and conduct continuous monitoring of vital signs.

(3) What are the implications of the data quality and individual characteristics on decision support systems like the EWS?

Decision support systems like the Early Warning System (EWS) rely on patient data compiled from intermittent vital signs records. However, advancements in technology now enable continuous monitoring of vital signs, providing a wealth of real-time data. As discussed in the literature chapter of this thesis, EWS was originally designed for intermittent vital signs monitoring. Incorporating continuous monitoring into the system has the potential to generate a massive amount of data, but it also offers significant advantages in identifying moments of patient deterioration that might otherwise go unnoticed.

Nevertheless, the EWS system has shown problems when performing with continuous monitoring systems and wireless devices. Usually, the systems send many false positive alarms, leading to nurse fatigue (Leenen et al., 2022). This fact indicates that the conventional EWS, designed for intermittent monitoring, may not be suitable for continuous monitoring scenarios with remote devices. To address this issue, researchers have attempted to develop customized EWS versions for continuous monitoring. These adaptations involve changing threshold values or employing advanced machine learning techniques to predict individualized approaches and reduce false negatives.

Lately, research has focused on the use of machine learning algorithms to identify specific patterns in the vital signs; however, it's crucial to keep in mind that, in the end, those approaches are not explanatory. Indeed, one of the primary goals of researchers is to comprehend the underlying relationships and factors influencing specific phenomena. In the case of EWS, it is crucial to investigate why the original threshold system may not perform optimally under continuous vital sign monitoring.

Understanding the limitations of the original threshold system is essential to refine and enhance EWS for continuous monitoring. Factors such as age, gender, body size, and other individual characteristics may influence vital signs and, subsequently, the appropriate threshold values for each patient.

The thesis provided evidence supporting the notion that heart rate distributions differed depending on age groups, with younger adults having significantly higher heart rates than older adults. EWS's current threshold system may not appropriately accommodate these differences (Leenen et al., 2023). For instance, following the results obtained in this study, a minor increase in heart rate could significantly impact the EWS score for younger patients (+ 6 bmp), whereas a substantial change is required for older patients (+ 22 bpm). This indicates that the threshold used in EWS is too low for younger patients and too high for older patients, compromising its accuracy for continuous monitoring.

This section concludes by emphasizing the significance of optimizing decision support systems such as EWS to align with continuous vital sign monitoring. Researchers can improve EWS performance and prepare the way for more effective and personalized patient care in the context of remote continuous vital sign monitoring by recognizing and addressing data quality issues, individual characteristics such as age-related variations in vital sign values.

6.3 Practical implications

a. Reflection of the study results to the healthcare practice:

The use of wearable patches for continuous vital sign monitoring has the potential to alleviate the strain on medical staff and reduce the time invested in individually recording vital signs. This technology opens up possibilities for remote monitoring of inside and outside patients. However, critical attention should be given to how to understand the shortcomings of this device.

The findings of this study have important implications for medical care, particularly in interpreting and understanding the results generated by continuous vital sign monitoring systems. The study revealed that patients with a higher BMI are at a higher risk of data loss in their vital signs. Moreover, the study also highlighted that vital signs' accuracy may be compromised during periods of high activity levels. By being cognizant of these factors, medical care teams can better evaluate the reliability and limitations of the continuous monitoring systems employed.

Additionally, it is critical to recognize that this big data trend in medical care is becoming more prevalent. So, in this regard, it is essential that healthcare professionals are well-versed in understanding biases and data loss in the recorded vital signs. Being aware of how a lack of data can influence decision-making and the algorithms used for data analysis Training healthcare providers in these areas will enhance their ability to interpret continuous vital sign data accurately and make informed clinical decisions.

Moreover, it requires a shift in how healthcare professionals interact with and interpret patient data. It calls for a deeper understanding of the technology and data science aspects, ensuring that healthcare professionals are not only proficient in using the wearable devices but also adept at leveraging the generated data to provide optimal patient care.

The development of new wearable devices for continuous vital sign monitoring should prioritize the needs of the medical staff and healthcare providers who will be using these devices to care for patients. Manufacturers must collaborate closely with medical professionals to ensure that the devices collect and provide relevant context along with the vital sign measurements.

As demonstrated in this study, the absence of references for activity levels was a significant drawback, as it hindered the ability to interpret the vital sign data accurately. Providing context for activity levels, such as clearly defining what each level represents in terms of physical exertion or movement, is crucial for healthcare providers to understand the implications of vital sign measurements during different activity states.

Finally, the last practical implication of the study is the need to redefine the threshold values used in decision support systems, especially in the context of continuous vital sign monitoring. Implementing group threshold values based on individual characteristics such as age can improve the accuracy of EWS, leading to more precise and timely alerts for healthcare providers. This can result in quicker responses to deteriorating patient conditions, reducing adverse events and hospital readmissions.

b. Implication of the use of ADS for Public Administration

The research findings presented in this thesis hold significant implications not only for healthcare practice but also for public administration and algorithmic decision support systems (ADS) across various sectors. The study sheds light on the central role of data quality and data loss within ADS, asserting that the data used in these systems holds equal importance to the results they provide, as it profoundly influences decision-making and outcomes. Additionally, it highlights the importance of acknowledgment based on individual characteristics rather than general standards.

A crucial aspect explored in this research is the investigation of whether the observed low data quality in ADS a random occurrence is or consistently lower for specific populations or under certain circumstances. The study demonstrates that data quality varies among distinct groups and situations, with certain conditions predisposing data loss and introducing biases within the system. As an example, the lack of information, or inconsistent data collection by ethnicity or gender has historically affects the results of ADM in sectors such education (Hu & Rangwala, 2020; Ramineni & Williamson, 2018) , healthcare (Jessani et al., 2022; Yi et al., 2022) and justice, (Berk et al., 2021; Hamilton, 2019) leading to recommendations that reinforced discrimination.

Further examination of this phenomenon in various administrative decision-making contexts is essential to fully grasping the limitations of its results. Ensuring the accuracy and high quality of data collected by ADS for every individual under its purview is paramount to ensuring equitable opportunities and outcomes for all (O'Neil, 2016). Recognizing and addressing these variations is essential to ensuring fairness and equal opportunities for all citizens, as stated.

Additionally, the present thesis suggests the use of more individualized standards over general standards in the Public Sector. ADS for service delivery should be applicable to a diverse population. Aggregate data lead to over generalization of certain results, and differentiation at least for demographic information is necessary to enable algorithmic fairness (Andrus & Villeneuve, 2022). While the present study finds that healthcare may showcase biological differences more directly, it is not rules out that other variables such as social roles and economic backgrounds may also play significant role.

As a last point, the study underscores the critical importance of investing in the advancement of public administrators' skills concerning the utilization and potential risks associated with Big Data and the algorithms they employ. A comprehensive analysis encompassing the input, process, and output of these algorithms can substantially empower public servants to identify and address false positives effectively . This becomes especially pertinent in light of research indicating that workers often exhibit a predisposition to unquestioningly accept outcomes produced by algorithms (Hitron et al., 2022; Robinette et al., 2017).

In specific domains like welfare aid, children's benefits, and migration, enhancing awareness regarding the inherent flaws and potential biases of algorithmic decision-making, in conjunction with human interpretation, holds the promise of yielding fewer errors and more efficient results (Kolkman, 2020). Although acknowledging the inevitability of imperfections in any system, the integration of human expertise and targeted training can lead to a more optimal solution. Such knowledge not only facilitates better decision-making but also empowers administrators to proactively identify and prevent errors before they escalate into larger issues.

In conclusion, this study advocates for a meticulous and discerning approach towards data quality, bias, and individualization Algorithmic Decision-Support systems in public administration. Embracing the insights gained from this study can lead to significant advancements in various sectors, fostering greater efficiency, fairness, and effectiveness in public service delivery.

6.4 Limitations of the study and future research lines:

The present study, while contributing valuable insights, does have several potential limitations that warrant consideration. These limitations encompass both the scope of the research and the data set employed, providing a foundation for future research endeavors in this domain.

Firstly, this study focuses on a restricted set of individual variables such as age, BMI, and gender. However, it is important to acknowledge that to effectively test hypotheses pertaining to differences in respiratory rate, the inclusion of additional variables such as height and weight is pertinent. The potential biological variations in lung size, which may lead to divergent respiratory rate baseline values, necessitate the incorporation of such variables. Furthermore, the influence of comorbidities and clinical history on vital sign patterns presents a compelling avenue for further investigation to ascertain how these factors impact continuous vital sign monitoring.

In regard to missing data, it is encouraged that future research include a detailed evaluation of the missing data problem and its potential effects on accuracy assessments in order to completely appraise the Heathdot's performance and the dependability of its measurements. Such efforts will aid in gaining a more complete knowledge of the device's capabilities and limits, thereby improving the clinical value and validity of its readings in bariatric patients and beyond.

Secondly, the data set employed for this thesis was derived from a specific population, which may not encompass the full spectrum of patient characteristics. Notably, the absence of representation for individuals with higher BMI levels (greater than 40) is a notable limitation. Similarly, the thesis is confined to the context of a particular country and the institutional settings surrounding it, potentially limiting the generalizability of the findings.

Finally, it is noteworthy to mention that the study covered a dataset of over 498,248 instances of heart rate and respiratory rate measurements, which were collected from a cohort of 384 patients. This indicates that the data could exhibit a certain level of clustering at the individual patient level. The conducted analysis utilized treat the data values as independent measurements. However, it is important to acknowledge that there are other variables, observed and unobserved, that can alter the values of vital signs.

Future research in this area should endeavor to address these limitations by incorporating participants from diverse countries, thus broadening the demographic representation. This approach will facilitate the inclusion of a more comprehensive range of individual variables, leading to a deeper understanding of the nuances and universalities of Algorithmic Decision-Support systems. Additionally, a promising direction for future research lies in the comparison of new thresholds and the development of novel early warning systems for continuous monitoring, thus enhancing the precision and effectiveness of such systems.

In conclusion, despite these acknowledged limitations, this research serves as an initial step towards assessing the quality of the data integrated into Algorithmic Decision-Support systems and comprehending the extent to which these systems accurately capture and represent individual characteristics. Building upon these foundational insights, future research endeavors can forge ahead in unraveling the intricate interplay between data quality, algorithmic models, and the personalized nature of healthcare decision-making.

7. Conclusions

The main objective of this thesis was to evaluate the quality of the data utilized in algorithmic decision-support systems and determine the extent to which these systems accurately capture and represent individual characteristics. The research focused on continuous monitoring of vital signs in hospitalized patients as a case study. The findings of this study have provided significant insights into optimizing data quality and enhancing our comprehension of the connections between vital signs and patient characteristics.

The examination of the data revealed noteworthy variations in data completeness across different recordings. These variations were predominantly influenced by activity levels, affecting 90% of vital signs data, as well as individual variables like BMI, which had an impact on 40% of heart rate data. Additionally, regarding individual characteristics and their connection with vital sign patterns, it was observed that younger patients exhibited a distinct heart rate distribution in comparison to older patients, with a significant difference of 18 bpm.

The implications of this study extend to the utilization of wearable devices and decision support systems, particularly Early Warning Systems (EWS), in healthcare. The results suggest that employing a one-size-fits-all approach may not yield optimal outcomes. Instead, adopting a more tailored and personalized approach to EWS implementation could potentially lead to improved patient outcomes.

Moreover, the problems and challenges discussed concerning algorithmic decision support in the continuous monitoring of vital signs can be relevant to other decision support systems utilized in public administration. In particular, the emphasis on data quality sheds light on possible unintended biases that tend to concentrate in certain circumstances rather than being randomly assigned. Additionally, evaluating different treatment approaches based on individual characteristics could be explored to ascertain whether such personalized strategies enhance the performance of algorithmic decision support.

In conclusion, this study advocates for a meticulous and discerning approach towards data quality, bias, and individualization in public Administration Decision Support Systems. The insights derived from this research have the potential to drive significant advancements in various sectors, fostering greater efficiency, fairness, and effectiveness in public service delivery. By embracing these findings and implementing responsible algorithmic decision support strategies, we can strive towards a more equitable and impactful approach to healthcare and public administration decision-making.

8. Acknowledgements

I wish to extend my appreciation to all the individuals who played an integral role in the realization of this study. Foremost, I am grateful for the support of the interdisciplinary thesis circle of Technology in Healthcare Transformations (THT), where the seeds of this project were nurtured through invaluable feedback from fellow students and researchers.

In addition, I am grateful to my supervisor, Dr. Caroline Fischer, whose guidance not only sparked my interest in healthcare research but also supported me throughout every step of this academic journey. Her guidance has been crucial in shaping the direction of this study.

Likewise, I would like to extend my sincere gratitude to Isala Hospital-Zwolle for providing the necessary infrastructure and crucial data that made this study possible. Special appreciation goes to Job Leenen for his continuous guidance throughout the entire process.

Lastly, my deepest appreciation goes to my husband, Javi, avi, who patiently supported me during late-night thesis discussions. His presence has been a pillar of strength during my master’s studies.

9. References

Alam, N., Hobbelink, E. L., van Tienhoven, A. J., van de Ven, P. M., Jansma, E. P., & Nanayakkara, P. W. B. (2014). The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review. In *Resuscitation* (Vol. 85, Issue 5). https://doi.org/10.1016/j.resuscitation.2014.01.013

An, X., Liu, Y., Zhao, Y., Lu, S., Stylios, G. K., & Liu, Q. (2022). Adaptive Motion Artifact Reduction in Wearable ECG Measurements Using Impedance Pneumography Signal. *Sensors*, *22*(15). https://doi.org/10.3390/s22155493

Avram, R., Tison, G. H., Aschbacher, K., Kuhar, P., Vittinghoff, E., Butzner, M., Runge, R., Wu, N., Pletcher, M. J., Marcus, G. M., & Olgin, J. (2019). Real-world heart rate norms in the Health eHeart study. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0134-9

Barocas, S., & Selbst, A. D. (2018). Big Data's Disparate Impact. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2477899

Bassi, D., Cabiddu, R., Mendes, R. G., Tossini, N., Arakelian, V. M., Caruso, F. C. R., Bonjorno Júnior, J. C., Arena, R., & Borghi-Silva, A. (2018). Effects of coexistence hypertension and type II diabetes on heart rate variability and cardiorespiratory fitness. *Arquivos Brasileiros de Cardiologia*, *111*(1). https://doi.org/10.5935/abc.20180105

Bennett, D. A. (2001). How can I deal with missing data in my study? In *Australian and New Zealand Journal of Public Health* (Vol. 25, Issue 5). https://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, *28*(7). https://doi.org/10.1080/10439463.2016.1253695

Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-0226-6

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods and Research*, *50*(1). https://doi.org/10.1177/0049124118782533

Bonnemeier, H., Wiegand, U. K. H., Brandes, A., Kluge, N., Katus, H. A., Richardt, G., & Potratz, J. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: Differing effects of aging and gender on heart rate variability. *Journal of Cardiovascular Electrophysiology*, *14*(8). https://doi.org/10.1046/j.1540-8167.2003.03078.x

Breteler, M. J. M., Huizinga, E., van Loon, K., Leenen, L. P. H., Dohmen, D. A. J., Kalkman, C. J., & Blokhuis, T. J. (2018). Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: A clinical validation study. *BMJ Open*, *8*(2). https://doi.org/10.1136/bmjopen-2017-020162

Buolamwini, J. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *. In *Proceedings of Machine Learning Research* (Vol. 81).

Cahan, E. M., Hernandez-Boussard, T., Thadaney-Israni, S., & Rubin, D. L. (2019). Putting the data before the algorithm in big data addressing personalized healthcare. In *npj Digital Medicine* (Vol. 2, Issue 1). https://doi.org/10.1038/s41746-019-0157-2

Capan, M., Hoover, S., Miller, K. E., Pal, C., Glasgow, J. M., Jackson, E. V., & Arnold, R. C. (2018). Data-driven approach to early warning score-based alert management. *BMJ Open Quality*, *7*(3). https://doi.org/10.1136/bmjoq-2017-000088

Cardona-Morrell, M., Prgomet, M., Turner, R. M., Nicholson, M., & Hillman, K. (2016). Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. In *International Journal of Clinical Practice* (Vol. 70, Issue 10). https://doi.org/10.1111/ijcp.12846

Chauhan, A. (2020). Towards the Systemic Review of Automated Decision-Making Systems. *Judicial Review*, *25*(4). https://doi.org/10.1080/10854681.2020.1871714

Chester, J. G., & Rudolph, J. L. (2011). Vital Signs in Older Patients: Age-Related Changes. In *Journal of the American Medical Directors Association* (Vol. 12, Issue 5). https://doi.org/10.1016/j.jamda.2010.04.009

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. Bin, Islam, K. R., Khan, M. S., Iqbal, A., Emadi, N. Al, Reaz, M. B. I., & Islam, M. T. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3010287

Craig, E. (2010). Applied Missing Data Analysis Craig K Enders. In T. Little (Ed.), *Australian & New Zealand Journal of Statistics* (Vol. 1, Issue 2). The Guilford Press. https://doi.org/10.1111/j.1467-842x.2012.00656.x

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. In *SpringerPlus* (Vol. 2, Issue 1). https://doi.org/10.1186/2193-1801-2-222

Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data and Society*, *3*(2). https://doi.org/10.1177/2053951716665128

Downey, C. L., Chapman, S., Randell, R., Brown, J. M., & Jayne, D. G. (2018). The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. In *International Journal of Nursing Studies* (Vol. 84). https://doi.org/10.1016/j.ijnurstu.2018.04.013

Downey, C. L., Tahir, W., Randell, R., Brown, J. M., & Jayne, D. G. (2017). Strengths and limitations of early warning scores: A systematic review and narrative synthesis. In *International Journal of Nursing Studies* (Vol. 76). https://doi.org/10.1016/j.ijnurstu.2017.09.003

Downey, C., Ng, S., Jayne, D., & Wong, D. (2019). Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: A clinical validation study from the TRaCINg trial. *BMJ Open*, *9*(8). https://doi.org/10.1136/bmjopen-2019-031150

Eccles, S. R., Subbe, C., Hancock, D., & Thomson, N. (2014). CREWS: Improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia. *Resuscitation*, *85*(1). https://doi.org/10.1016/j.resuscitation.2013.08.277

Eubanks, V. (2017). *Automating Inequality: How high-tech tools profile, police, and punish the poor*.

Fang, A. H. Sen, Lim, W. T., & Balakrishnan, T. (2020). Early warning score validation methodologies and performance metrics: a systematic review. *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-01144-8

Fararouei, M., Marzban, M., & Shahraki, G. (2017). Completeness of cancer registry data in a small Iranian province: A capture-recapture approach. *Health Information Management Journal*, *46*(2). https://doi.org/10.1177/1833358316668605

Fleming, S., Thompson, M., Stevens, R., Heneghan, C., Plüddemann, A., MacOnochie, I., Tarassenko, L., & Mant, D. (2011). Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *The Lancet*, *377*(9770). https://doi.org/10.1016/S0140-6736(10)62226-X

Gijsberts, C. M., Groenewegen, K. A., Hoefer, I. E., Eijkemans, M. J. C., Asselbergs, F. W., Anderson, T. J., Britton, A. R., Dekker, J. M., Engström, G., Evans, G. W., De Graaf, J., Grobbee, D. E., Hedblad, B., Holewijn, S., Ikeda, A., Kitagawa, K., Kitamura, A., De Kleijn, D. P. V., Lonn, E. M., … Den Ruijter, H. M. (2015). Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. *PLoS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.0132321

Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. In *Computers in Biology and Medicine* (Vol. 127). https://doi.org/10.1016/j.compbiomed.2020.104065

Hamilton, M. (2019). The sexist algorithm. *Behavioral Sciences and the Law*, *37*(2). https://doi.org/10.1002/bsl.2406

Hitron, T., Megidish, B., Todress, E., Morag, N., & Erel, H. (2022). AI bias in Human-Robot Interaction: An evaluation of the Risk in Gender Biased Robots. *RO-MAN 2022 - 31st IEEE International Conference on Robot and Human Interactive Communication: Social, Asocial, and Antisocial Robots*. https://doi.org/10.1109/RO-MAN53752.2022.9900673

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2). https://doi.org/10.1080/10919392.2015.1125187

Holten Mandøller, N., Shklovski, I., & Hildebrandt, T. T. (2020). Shifting Concepts of Value: Designing Algorithmic Decision-Support Systems for Public Services. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3419249.3420149

Hu, Q., & Rangwala, H. (2020). Towards Fair Educational Data Mining: A Case Study on Detecting At-risk Students. *Proceedings of The 13th International Conference on Educational Data Mining, Edm*.

Jacobs, F., Scheerhoorn, J., Mestrom, E., van der Stam, J., Arthur Bouwman, R., & Nienhuijs, S. (2021). Reliability of heart rate and respiration rate measurements with a wireless accelerometer in postbariatric recovery. *PLoS ONE*, *16*(4 April 2021). https://doi.org/10.1371/journal.pone.0247903

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Medical Research Methodology*, *17*(1). https://doi.org/10.1186/s12874-017-0442-1

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. In *Government Information Quarterly* (Vol. 33, Issue 3). https://doi.org/10.1016/j.giq.2016.08.011

Jensen, M. T., Marott, J. L., Lange, P., Vestbo, J., Schnohr, P., Nielsen, O. W., Jensen, J. S., & Jensen, G. B. (2013). Resting heart rate is a predictor of mortality in COPD. *European Respiratory Journal*, *42*(2). https://doi.org/10.1183/09031936.00072212

Jessani, N. S., Williamson, R. T., Choonara, S., Gautier, L., Hoe, C., Jafar, S. K., Khalid, A. F., Rodríguez Salas, I., Turcotte-Tremblay, A. M., & Rodríguez, D. C. (2022). Evidence attack in public health: Diverse actors' experiences with translating controversial or misrepresented evidence in health policy and systems research. *Global Public Health*, *17*(11). https://doi.org/10.1080/17441692.2021.2020319

Joost Schellevis. (2021, February 1). *Fraude opsporen of gevaar van discriminatie? Gemeenten gebruiken "slimme" algoritmes*. NOS News. https://nos.nl/artikel/2366864-fraude-opsporen-of-gevaar-van-discriminatie-gemeenten-gebruiken-slimme-algoritmes

Kant, N., Peters, G. M., Voorthuis, B. J., Groothuis-Oudshoorn, C. G. M., Koning, M. V., Witteman, B. P. L., Rinia-Feenstra, M., & Doggen, C. J. M. (2022). Continuous vital sign monitoring using a wearable patch sensor in obese patients: a validation study in a clinical setting. *Journal of Clinical Monitoring and Computing*, *36*(5). https://doi.org/10.1007/s10877-021-00785-y

Keddell, E. (2019). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Sciences*, *8*(10). https://doi.org/10.3390/socsci8100281

Keller Matthew, Hrrison-Smith Brandon, Patil Chetan, & Arefin Mohammed Shahriar. (2022). Skin colour affects oxygen-sensor accuracy Expert insight into current research. *Nature*, *610*(7932).

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., … Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, *172*(5). https://doi.org/10.1016/j.cell.2018.02.010

Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: "Garbage in – garbage out." In *Health Information Management Journal* (Vol. 47, Issue 3). https://doi.org/10.1177/1833358318774357

Koenig, J., & Thayer, J. F. (2016). Sex differences in healthy human heart rate variability: A meta-analysis. In *Neuroscience and Biobehavioral Reviews* (Vol. 64). https://doi.org/10.1016/j.neubiorev.2016.03.007

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, *37*(3). https://doi.org/10.1016/j.giq.2020.101488

König, P. D., & Wenzelburger, G. (2021). The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society*, *67*. https://doi.org/10.1016/j.techsoc.2021.101688

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6). https://doi.org/10.1016/j.telpol.2020.101976

Langsted, S. N. D., Dynesen, J. J., Liesanth, J. Y., Jessen, M. K., MacKenhauer, J., Ahrensberg, J., & Kirkegaard, H. (2020). How comorbidities impact Early Warning Score as a predictor of 7-day mortality. *European Journal of Emergency Medicine*, *27*(2). https://doi.org/10.1097/MEJ.0000000000000633

le Lagadec, M. D., & Dwyer, T. (2017). Scoping review: The use of early warning systems for the identification of in-hospital patients at risk of deterioration. In *Australian Critical Care* (Vol. 30, Issue 4). https://doi.org/10.1016/j.aucc.2016.10.003

Lee, J. H., & Huber, J. C. (2021). Evaluation of multiple imputation with large proportions of missing data: How much is too much? *Iranian Journal of Public Health*, *50*(7).

Leenen, J. P. L., Leerentveld, C., van Dijk, J. D., van Westreenen, H. L., Schoonhoven, L., & Patijn, G. A. (2020). Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: Systematic review. In *Journal of Medical Internet Research* (Vol. 22, Issue 6). https://doi.org/10.2196/18636

Leenen, J. P. L., Rasing, H. J. M., Kalkman, C. J., Schoonhoven, L., & Patijn, G. A. (2023). MULTIMEDIA APPENDIX 4: Thresholds of the partial Early Warning (D-EWS) scores. Process Evaluation of a Wireless Wearable Continuous Vital Signs Monitoring Intervention in 2 General Hospital Wards: Mixed Methods Study. *JMIR Nursing*, *6*, e44061. https://doi.org/10.2196/44061

Leenen, J. P. L., Rasing, H. J. M., Van Dijk, J. D., Kalkman, C. J., Schoonhoven, L., & Patijn, G. A. (2022). Feasibility of wireless continuous monitoring of vital signs without using alarms on a general surgical ward: A mixed methods study. *PLoS ONE*, *17*(3 March). https://doi.org/10.1371/journal.pone.0265435

Leuvan, C. H. V., & Mitchell, I. (2008). Missed opportunities? An observational study of vital sign measurements. *Critical Care and Resuscitation : Journal of the Australasian Academy of Critical Care Medicine*, *10*(2).

Li, C. (2013). Little's test of missing completely at random. *Stata Journal*, *13*(4). https://doi.org/10.1177/1536867x1301300407

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404). https://doi.org/10.1080/01621459.1988.10478722

Littleton, S. W. (2012). Impact of obesity on respiratory function. In *Respirology* (Vol. 17, Issue 1). https://doi.org/10.1111/j.1440-1843.2011.02096.x

Ljunggren, M., Castrén, M., Nordberg, M., & Kurland, L. (2016). The association between vital signs and mortality in a retrospective cohort study of an unselected emergency department population. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *24*(1). https://doi.org/10.1186/s13049-016-0213-8

Lomauro, A., & Aliverti, A. (2018). Sex differences in respiratory function. *Breathe*, *14*(2). https://doi.org/10.1183/20734735.000318

Loyola-Gonzalez, O. (2019). Black-box vs. White-Box: Understanding their advantages and weaknesses from a practical point of view. In *IEEE Access* (Vol. 7). https://doi.org/10.1109/ACCESS.2019.2949286

Manta, C., Jain, S. S., Coravos, A., Mendelsohn, D., & Izmailova, E. S. (2020). An Evaluation of Biometric Monitoring Technologies for Vital Signs in the Era of COVID-19. In *Clinical and Translational Science* (Vol. 13, Issue 6). https://doi.org/10.1111/cts.12874

Miron, M., Tolan, S., Gómez, E., & Castillo, C. (2021). Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial Intelligence and Law*, *29*(2). https://doi.org/10.1007/s10506-020-09268-y

Monarcha-Matlak, A. (2021). Automated decision-making in public administration. *Procedia Computer Science*, *192*. https://doi.org/10.1016/j.procs.2021.08.215

Nagarajah, S., Krzyzanowska, M. K., & Murphy, T. (2022). Early Warning Scores and Their Application in the Inpatient Oncology Settings. *JCO Oncology Practice*, *18*(6). https://doi.org/10.1200/op.21.00532

Nilsson, M., Jordan, A., Turnpenny, J., Hertin, J., Nykvist, B., & Russel, D. (2008). The use and non-use of policy appraisal tools in public policy making: An analysis of three European countries and the European Union. *Policy Sciences*, *41*(4). https://doi.org/10.1007/s11077-008-9071-1

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464). https://doi.org/10.1126/science.aax2342

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'experimental' proportionality. *Information and Communications Technology Law*, *27*(2). https://doi.org/10.1080/13600834.2018.1458455

Paulus, J. K., Wessler, B. S., Lundquist, C. M., & Kent, D. M. (2018). Effects of Race Are Rarely Included in Clinical Prediction Models for Cardiovascular Disease. *Journal of General Internal Medicine*, *33*(9). https://doi.org/10.1007/s11606-018-4475-x

Petersen, J. A. (2018). Early warning score challenges and opportunities in the care of deteriorating patients. *Danish Medical Journal*, *65*(2).

Philips Electronics Nederland B.V. (2020). *Heathdot Technical Specifications*.

Pinto, E. (2007). Blood pressure and ageing. In *Postgraduate Medical Journal* (Vol. 83, Issue 976). https://doi.org/10.1136/pgmj.2006.048371

Posthuma, L. M., Downey, C., Visscher, M. J., Ghazali, D. A., Joshi, M., Ashrafian, H., Khan, S., Darzi, A., Goldstone, J., & Preckel, B. (2020). Remote wireless vital signs monitoring on the ward for early detection of deteriorating patients: A case series. In *International Journal of Nursing Studies* (Vol. 104). https://doi.org/10.1016/j.ijnurstu.2019.103515

Posthuma, L. M., Visscher, M. J., Lirk, P. B., van Dijkum, E. J. M. N., Hollmann, M. W., & Preckel, B. (2020). Insights into postoperative respiration by using continuous wireless monitoring of respiratory rate on the postoperative ward: a cohort study. *Journal of Clinical Monitoring and Computing*, *34*(6). https://doi.org/10.1007/s10877-019-00419-4

Ramineni, C., & Williamson, D. (2018). Understanding Mean Score Differences Between the e-rater® Automated Scoring Engine and Humans for Demographically Based Groups in the GRE® General Test. *ETS Research Report Series*, *2018*(1). https://doi.org/10.1002/ets2.12192

Robinette, P., Howard, A. M., & Wagner, A. R. (2017). Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations. *IEEE Transactions on Human-Machine Systems*, *47*(4). https://doi.org/10.1109/THMS.2017.2648849

Santos, M. D., Roman, C., Pimentel, M. A. F., Vollam, S., Areia, C., Young, L., Watkinson, P., & Tarassenko, L. (2021). A Real-Time Wearable System for Monitoring Vital Signs of COVID-19 Patients in a Hospital Setting. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.630273

Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (2021). A Framework of High-Stakes Algorithmic Decision-Making for the Public Sector Developed through a Case Study of Child-Welfare. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). https://doi.org/10.1145/3476089

Shcherbina, A., Mikael Mattsson, C., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Wheeler, M. T., & Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, *7*(2). https://doi.org/10.3390/jpm7020003

Shindo, K., Morishima, Y., Suwa, Y., Fukao, T., Kurita, T., Satake, A., Tsuchiya, M., Ichinose, Y., Hata, T., Koh, K., Nagasaka, T., & Takiyama, Y. (2021). Age-related changes in blood pressure and heart rates of patients with Parkinson's disease. *Journal of Clinical Hypertension*, *23*(1). https://doi.org/10.1111/jch.14096

Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E., & Featherstone, P. I. (2013). The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*, *84*(4). https://doi.org/10.1016/j.resuscitation.2012.12.016

Smith, M. E. B., Chiovaro, J. C., O'Neil, M., Kansagara, D., Quiñones, A. R., Freeman, M., Motu'apuaka, M. L., & Slatore, C. G. (2014). Early warning system scores for clinical deterioration in hospitalized patients: A systematic review. In *Annals of the American Thoracic Society* (Vol. 11, Issue 9). https://doi.org/10.1513/AnnalsATS.201403-102OC

Subbe, C. P., & Bramley, R. (2022). Digital NEWS? How to amplify the benefits of NEWS in a digital healthcare system. *Clinical Medicine, Journal of the Royal College of Physicians of London*, *22*(6). https://doi.org/10.7861/clinmed.2022-0349

Takayama, A., Nagamine, T., & Kotani, K. (2019). Aging is independently associated with an increasing normal respiratory rate among an older adult population in a clinical setting: A cross-sectional study. *Geriatrics and Gerontology International*, *19*(11). https://doi.org/10.1111/ggi.13788

van der Stam, J. A., Mestrom, E. H. J., Nienhuijs, S. W., de Hingh, I. H. J. T., Boer, A. K., van Riel, N. A. W., de Groot, K. T. J., Verhaegh, W., Scharnhorst, V., & Bouwman, R. A. (2022). A wearable patch based remote early warning score (REWS) in major abdominal cancer surgery patients. *European Journal of Surgical Oncology*. https://doi.org/10.1016/j.ejso.2022.08.034

van der Stam, J. A., Mestrom, E. H. J., Scheerhoorn, J., Jacobs, F., de Hingh, I. H. J. T., van Riel, N. A. W., Boer, A. K., Scharnhorst, V., Nienhuijs, S. W., & Bouwman, R. A. (2022). Accuracy of vital parameters measured by a wearable patch following major abdominal cancer surgery. *European Journal of Surgical Oncology*, *48*(4). https://doi.org/10.1016/j.ejso.2021.10.034

van Rossum, M. C., Vlaskamp, L. B., Posthuma, L. M., Visscher, M. J., Breteler, M. J. M., Hermens, H. J., Kalkman, C. J., & Preckel, B. (2022). Adaptive threshold-based alarm strategies for continuous vital signs monitoring. *Journal of Clinical Monitoring and Computing*, *36*(2). https://doi.org/10.1007/s10877-021-00666-4

Weenk, M., Koeneman, M., van de Belt, T. H., Engelen, L. J. L. P. G., van Goor, H., & Bredie, S. J. H. (2019). Wireless and continuous monitoring of vital signs in patients at the general ward. *Resuscitation*, *136*. https://doi.org/10.1016/j.resuscitation.2019.01.017

Wei, Y. C., George, N. I., Chang, C. W., & Hicks, K. A. (2017). Assessing sex differences in the risk of cardiovascular disease and mortality per increment in systolic blood pressure: A systematic review and meta-analysis of follow-up studies in the United States. *PLoS ONE*, *12*(1). https://doi.org/10.1371/journal.pone.0170218

Wells, C. I., Xu, W., Penfold, J. A., Keane, C., Gharibans, A. A., Bissett, I. P., & O'Grady, G. (2022). Wearable devices to monitor recovery after abdominal surgery: Scoping review. In *BJS Open* (Vol. 6, Issue 2). https://doi.org/10.1093/bjsopen/zrac031

Wysocki, O., Davies, J. K., Vigo, M., Armstrong, A. C., Landers, D., Lee, R., & Freitas, A. (2023). Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making. *Artificial Intelligence*, *316*. https://doi.org/10.1016/j.artint.2022.103839

Yeragani, V. K., Sobolewski, E., Kay, J., Jampala, V. C., & Igel, G. (1997). Effect of age on long-term heart rate variability. *Cardiovascular Research*, *35*(1). https://doi.org/10.1016/S0008-6363(97)00107-7

Yi, S. S., Kwon, S. C., Suss, R., Đoàn, L. N., John, I., Islam, N. S., & Trinh-Shevrin, C. (2022). The Mutually Reinforcing Cycle Of Poor Data Quality And Racialized Stereotypes That Shapes Asian American Health. *Health Affairs*, *41*(2), 296–303. https://doi.org/10.1377/hlthaff.2021.01417

Zarsky, T. (2016). The Trouble with Algorithmic Decisions. *Science, Technology, & Human Values*, *41*(1). https://doi.org/10.1177/0162243915605575

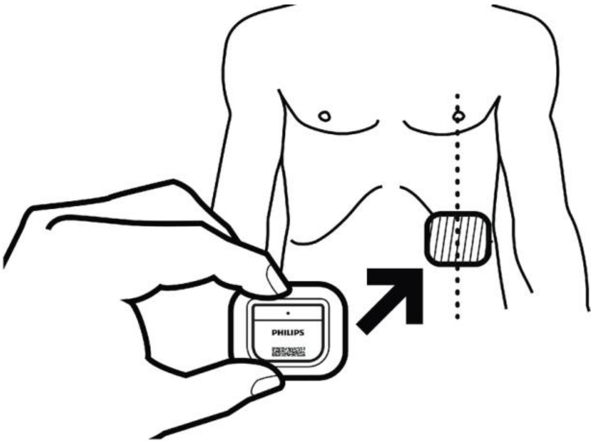
Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, *18*(5). https://doi.org/10.1177/1477370819876762

Zhang, J. (2007). Effect of Age and Sex on Heart Rate Variability in Healthy Subjects. *Journal of Manipulative and Physiological Therapeutics*, *30*(5). https://doi.org/10.1016/j.jmpt.2007.04.001

Zhao, Z., Wu, J., Li, T., Sun, C., Yan, R., & Chen, X. (2021). Challenges and Opportunities of AI-Enabled Monitoring, Diagnosis & Prognosis: A Review. In *Chinese Journal of Mechanical Engineering (English Edition)* (Vol. 34, Issue 1). https://doi.org/10.1186/s10033-021-00570-7

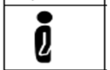
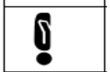
10. Appendix

Appendix 1. Collocation of the Healthdot in the patient



Note: The wearable patch is attached in lower left rib. From “Reliability of heart rate and respiration rate measurements with a wireless accelerometer in postbariatric recovery” by Philips Electronic Nederland BV(2020) (doi: https://doi.org/10.1371/journal.pone.0247903.g001). Under a CC BY license, original copyright 2020.

Appendix 2. Position of the patient measured by the Healthdot

Pictogram	Position of the patient
	Forward
	Backward
	Upright
	Other way around
	Lying left
	Supine position
	Lying right
	Lying on belly

Adpated from “Heathot Technical Sheet” by Philips Electronics Nederland B.V, (2020).