Corvinus University of Budapest

Doctoral School of Business and Management

Methodological challenges in the evidence synthesis of health outcomes of digital health technologies

PhD. Dissertation

Supervisor: Dr. Zsombor Zrubka, MD, MBA, Ph.D.

Hossein Motahari Nezhad

Budapest, 2023

Hossein Motahari Nezhad

Corvinus University of Budapest

Department of Health Economics

Supervisor:

Dr. Zrubka Zsombor, MD, MBA, PhD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to the following individuals and institutions who have played a significant role in the completion of my study:

First and foremost, I extend my sincere appreciation to my supervisor, Dr. Zrubka Zsombor, for his exceptional guidance, invaluable insights, and unwavering support throughout my research journey. His expertise, dedication, and encouragement have been instrumental in shaping this work and expanding my academic horizons. I am deeply grateful to Corvinus University of Budapest, the Doctoral School of Business and Management, and the Department of Health Economics for providing me with the resources, facilities, and academic environment necessary for carrying out this career. Their commitment to excellence in education and research has been a constant source of inspiration. I would like to acknowledge the Doctoral Office for their invaluable assistance and support during my time as a PhD student. Their responses, administrative guidance, and willingness to address any concerns have been truly commendable, and I am grateful for their efforts in ensuring a smooth progression throughout the doctoral program. I would also like to extend my sincere appreciation to Stipendium Hungaricum for providing me with the opportunity to pursue my studies in Hungary.

Finally, I would like to express my deepest gratitude to HECON at Obuda University, Prof. Laszlo Gulácsi and Prof. Márta Péntek, my family and friends for their encouragement and support throughout this journey. Their understanding, patience, and belief in me have been my pillars of strength.

To everyone mentioned above, and to all those who have contributed in ways big and small, I extend my heartfelt thanks. Your support and guidance have been invaluable, and I am truly grateful for the impact you have had on my academic and personal development.

LIST OF ABBREVIATIONS

FDA: Food and Drug Administration

IT: Information technology

ICD: Implantable cardiac defibrillator

ICDs: Implantable cardiac defibrillators

DBM: Digital biomarker

DBMs: Digital biomarkers

CAGR: Compound annual growth rate

SRs: Systematic reviews

SR: Systematic review

MA: Meta-analysis

MAs: Meta-analyses

PICO: Population, Intervention, Comparator, Outcome

WHO: World Health Organization

ICD: International Classification of Diseases

ICHI: International Classification of Health Interventions

ICF: International Classification of Functioning, Disability and Health

AMSTAR: A measurement tool for assessment of multiple systematic reviews

GRADE: Grading of Recommendations Assessment, Development and Evaluation

RCT: Randomized controlled trial

DHTs: Digital health technologies

HTA: Health technology assessment

ESF: The Evidence Standards Framework

NICE: The National Institute for Health and Care Excellence

ICT: Information and communication technology

mHealth: Mobile health

NHS: National Health Service

TAM: Technology Acceptance Model

UTAUT: The Unified Theory of Acceptance and Use of Technology

COPD: Chronic Obstructive Pulmonary Disease

CIED: Cardiac Implantable Electronic Device

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

BMI: Body mass index

T1D: Type 1 diabetes

SLR: Systematic literature review

EBM: Evidence-based medicine

EMA: European Medicines Agency

ROSC: Return of spontaneous circulation

T2DM: Type 2 diabetes mellitus

MVPA: Moderate to vigorous physical activity

HF: Heart failure

I. INTRODUCTION

With the advent of medical technology and new drugs, medical practice and prognosis of diseases with high mortality rates have undergone a remarkable transformation (Holmes and Wood, 2006). Without the contributions of the medical device and pharmaceutical industries, healthcare would be inconceivable, although these two professions are worlds apart (MedTech, 2015). Drugs are compounds of chemical origin that are designed to interact dynamically in some way with the body's metabolic or immune systems (MedTech, 2015). Contrary to drugs, the vast majority of devices function mechanically and have no visible effects on the human body. Furthermore, some devices improve the information about patients or treatments, without direct biological action. According to the Food and Drug Administration (FDA), health information technology (IT), wearable devices, mobile health, and telehealth all fall under the term "digital health" (FDA, 2020). However, the use of both drugs and technologies in clinical practice carries the potential for adverse effects. As with any drug, side effects can range from minor inconveniences to potentially fatal reactions. The hazards that drugs pose to the body are of a different nature and magnitude than those posed by medical devices (MedTech, 2015).

While there are altogether less than 4300 approved drug molecules in the world (drugbank, 2022), the number of digital health solutions is increasing at a breakneck pace; there are already over 300,000 health apps accessible, with another 200 being developed every day (IQVIA, 2017). While the efficacy and safety of medicinal products is demonstrated in well-established development regulations, regulations for medical device development are evolving, including those of digital health. It is challenge to keep up with the pace and diversity of development in the device segment, while patients and health care professionals demand the same quality of evidence about their efficacy and safety.

This thesis focus on a special class of digital medical devices: digital biomarkers. *"Digital biomarkers are objective, measurable, physiological, and behavioural parameters collected using wearable, portable, implantable, or digestible digital devices"* (Babrak *et al.*, 2019). For instance, implantable cardiac defibrillators (ICDs) (Mahdi Abid *et al.*, 2022) or physical activity trackers (Ringeval *et al.*, 2020) can be considered as digital biomarkers. With the transition to digitization of healthcare, the term "digital biomarker" is increasingly being used to describe a wide range of measures (Vasudevan *et al.*, 2022). Digital biomarkers (DBMs)

may assist enhance patient outcomes by boosting diagnostic precision, elevating the quality of treatment choices, and decreasing the frequency with which clinical mistakes occur (Insel, 2017; Lipsmeier *et al.*, 2018; Shin *et al.*, 2018). Patient visits may be reduced while still achieving high quality outcomes thanks to digital biomarkers (Guthrie *et al.*, 2019). Digital biomarkers may enhance the accuracy of diagnosis and therapy by measuring clinical data remotely and continuously (Robb, McInnes and Califf, 2016; Lipsmeier *et al.*, 2018). For instance, for people with diabetes, the use of sensors to monitor their blood glucose levels in real time has the potential to inform individualized insulin dosing and provide advance warning of dangerously low readings (Vettoretti *et al.*, 2020). Demand for DBMs is expected to grow at a compound annual growth rate (CAGR) of 40.4% between 2019 and 2025, with global revenue reaching $5.64 billion by the end of the forecast period (Meister, Deiters and Becker, 2016; BisResearch, 2020). Overall, digital biomarkers play an important role in precision medicine (Jeong, Bychkov and Searson, 2019), can reduce the occurrence of clinical errors, increase the precision of diagnostic techniques, and support patient-specific treatment decisions (Insel, 2017). Depending on the interaction of digital biomarkers with the human body, we divide these digital instruments into two categories: direct and indirect. For example, defibrillators that regulate heart rhythm may be called direct digital biomarkers that directly impact physiological parameters without the interference of a physician or the patient. The utility of such instruments depends mainly on the technology involved. However, indirect digital biomarkers such as activity trackers just capture behavioural data such as heart rhythm or step count. To influence health, the information from indirect biomarkers has to be processed and acted upon by patients or health care professionals. Therefore, their utility depends on additional human factors. The diversity of applied technologies and complexity of the involved human-machine interaction poses unique challenges when demonstrating the clinical effectiveness and safety of digital biomarkers.

As a mainstay of evidence synthesis in medicine and the medical industries, systematic reviews (SRs) and meta-analyses (MAs) have gained prominence since the 1970s. They provide evidence-based information to inform decision making in medicine (Li *et al.*, 2021). Evidence-based medicine and clinical guideline development require rigorous review (Rabar *et al.*, 2012; Goff *et al.*, 2014). SRs and MAs serve as the foundation for clinical decision making (Gopalakrishnan and Ganeshkumar, 2013). Clinical decision-making relies heavily on the results of systematic reviews and meta-analyses. However, if certain procedures and criteria are not adhered to, systematic reviews and meta-analyses will provide poor quality results that

may have irreversible adverse repercussions for patients (Yuan and Hunt, 2009). It is highly recommended by the Cochrane Handbook (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a) for systematic reviews to follow some specific procedures during the study stages to avoid biases, and maintain high methodological quality. One of the Cochrane Handbook's suggestions is to formulate a complete research question while searching for systematic reviews. In this context, the PICO style, which stands for population, intervention, comparator, and outcome, has been advocated for clinical research. Formulating a research question requires a centralized and integrated system capable of categorizing the PICO of individual studies in systematic reviews using approved methodologies. The World Health Organization has proposed three proven techniques to classify population, intervention, and outcome in clinical research: ICD (International Classification of Diseases) (WHO, 2020a), ICHI (International Classification of Health Interventions) (World Health Organization, 2020), and ICF (International Classification of Functioning, Disability and Health) (WHO, 2017a) tools.

AMSTAR-2 (Shea *et al.*, 2017b), is one of the validated tool for researchers to evaluate the methodological quality of systematic reviews and meta-analyses. The "effect size" or "effect estimate" is a statistical finding that is the pooled result of meta-analyses that integrate the findings of individual related studies on a research issue (Sullivan and Feinn, 2012). The vast majority of clinical decisions are dependent on the magnitude and direction of effect sizes revealed in meta-analyses (Aarts, van den Akker and Winkens, 2014). One of the worldwide credible techniques for analysing the validity and quality of the reported effect sizes in meta-analyses is GRADE (Kumar and Taggarsi, 2021), which assesses the quality of the calculated effect sizes, also known as evidence quality. Prior to making any medical decisions, the Cochran Handbook recommends assessing the quality of the evidence (Lefebvre *et al.*, 2019). In recent years, scientists and funders in the medical sciences have expressed concern about what has been called the replication and reliability crisis in clinical research (Barch and Yarkoni, 2013). Systematic reviews and meta-analyses, which are regarded as the apex of the traditional hierarchy of study quality, are often erroneous (Ioannidis, 2016). Concerns about insufficient sample size and lack of statistical power have received much attention in both primary studies and meta-analyses (Brok *et al.*, 2008, 2009; Thorlund *et al.*, 2011). Type II errors are more likely in randomized controlled trials (RCTs) with small sample sizes, emphasizing the need for optimum sample sizes to enhance statistical power (Sjögren and Hedström, 2010). If a meta-analysis was unable to incorporate and exclude some relevant

studies, the effect estimates of the meta-analysis may not be accurate and may be inflated or understated, a phenomenon, which called publication bias. Publication bias may have a detrimental impact on the validity of effect size findings from meta-analyses (Kicinski, Springate and Kontopantelis, 2015).

Due to the diversity of technologies, the requirement that all included studies in systematic reviews target the same intervention mechanism, disease area, and quantitative outcomes may hinder digital health research (Guo *et al.*, 2020). The ability of researchers to provide evidence for digital health treatments is severely limited by traditional methods. However, the evaluation of digital health solutions has been highlighted as needing improvement and considered a major barrier to widespread adoption (Moxey *et al.*, 2010; O'Sullivan *et al.*, 2014; Shaw *et al.*, 2018). Despite these challenges, in recent years, an increasing number of systematic reviews and meta-analyses of digital biomarkers have been published and the relevance of digital biomarkers in clinical research is predicted to increase rapidly in the next years. A search of PubMed on 20 August 2022 revealed, for instance, that the number of published systematic reviews and meta-analyses in digital biomarkers had risen from 76 in 2016 to 264 in 2021. Particularly, the fast growth of systematic reviews in digital health (Ibrahim *et al.*, 2022) has been attributed, at least in part, to the increased output of systematic reviews in digital biomarkers. As previously noted, any methodological flaws that occur during the conduct of systematic reviews and meta-analyses might invalidate the results, resulting in irreversible damage to patients. In terms of rules, standards, etc., there are also distinctions between digital and non-digital or pharmaceutical products, as described in the first section of the introduction. No research has, to our knowledge, evaluated the methodological issues of digital biomarker-based studies and compared them to non-digital biomarkers. In this dissertation, we evaluated key methodological issues in digital biomarker research and compared them to non-digital biomarkers and pharmaceuticals. The results of this thesis will have significant implications for researchers and managers in the digital health technologies industry and digital biomarkers. Through comprehensive discussions on classification systems and study power within this field, the thesis will provide valuable insights. Additionally, by comparing digital biomarkers with non-digital biomarkers and pharmaceuticals in terms of evidence generation, the thesis will present both positive and negative aspects of evidence creation in the field. These insights will assist researchers, health policy makers, and managers in digital health technology and pharmaceutical industries in improving the quality of evidence and establishing more effective regulations to address existing gaps.

This thesis is divided into twelve chapters. The first chapter (Chapter I) is about the general introduction to the work. In this chapter, some concepts related to the thesis, the reasons and the importance of the subject are presented. In the second chapter (Chapter II), the relevant literature is listed. The hypotheses of the thesis are stated in the third chapter (chapter III). From the fourth chapter (IV) to the eighth chapter (VIII), the main part of the thesis is to test the hypotheses. Each chapter of the thesis includes introduction, methods, results, discussion and conclusion. Chapter IV deals with the digital biomarker classification system. In chapter V, the statistical power of studies of direct digital biomarkers was compared with that of indirect digital biomarkers. Chapter VI compared the methodological quality of systematic reviews of digital biomarkers with that of non-digital biomarkers or drugs. In chapter VII, the quality of evidence of digital biomarker based-meta-analyses was compared with that of non-digital biomarkers or pharmaceuticals. In the final chapter (VIII) evaluating the hypotheses, publication bias in meta-analyses of digital biomarkers was compared with that of non-digital biomarkers. The conclusions and the expected practical implications are explained in the chapter IX. I have listed my own publications in Chapter X. Chapter XI contains the references, and finally the appendices are listed at the end of the thesis.

II. LITERATURE REVIEW

II.1 Classification systems in medical research

II.1.1 The need for better classification systems of digital health interventions

The World Health Organization (WHO) defines "digital health" as a catch-all term for a variety of solutions that use digital technologies to address health needs, including eHealth, mHealth as well as emerging areas such as the use of advanced computer science in "Big Data," genomics, and artificial intelligence (*WHO Guideline*, 1980). Apps, computer programs, and web-based services are just a few examples of the many types of digital health technologies (DHTs) that may help individuals or the whole health care system. For healthcare systems on the lookout for more efficient, cost-effective, patient-cantered methods of providing treatment at scale, DHTs bring both potential and problems (Ferretti, Ronchi and Vayena, 2019). Due to the relative lack of standardized evaluation by healthcare systems, many digital health products are being offered to the market with little proof of their efficacy. Some have claimed that the health technology assessment (HTA) of digital health technologies (DHTs) should employ a different degree and kind of evidence than other treatments because of the apparent contradiction between typical HTA methodologies and the fast pace of DHT development (The Lancet, 2018). Therefore, it is necessary to review the relevant literature to determine what strategies and standards are introduced to categorize digital and non-digital health domains.

II.1.2 Categorisation systems for digital health interventions

II.1.2.1 The Evidence Standards Framework (ESF) of the National Institute for Health and Care Excellence (NICE)

An evidence standards framework (ESF) for digital healthcare technologies (DHTs) was developed in 2018 (NICE, 2019) by the National Institute for Health and Care Excellence (NICE) of the United Kingdom in partnership with Public Health England, the National Health Service (NHS England), and other organizations. The ESF was developed to provide a consistent method for developers and commissioners about the levels of evidence that are necessary for the clinical and economic evaluation of DHTs by health and care systems. A customized taxonomy of ten functional categories was created that encompasses the roles of most of the DHTs most commonly commissioned by the UK health and care system. The ten

functional categories are grouped into "tiers" based on the potential risk associated with each category (figure 1).

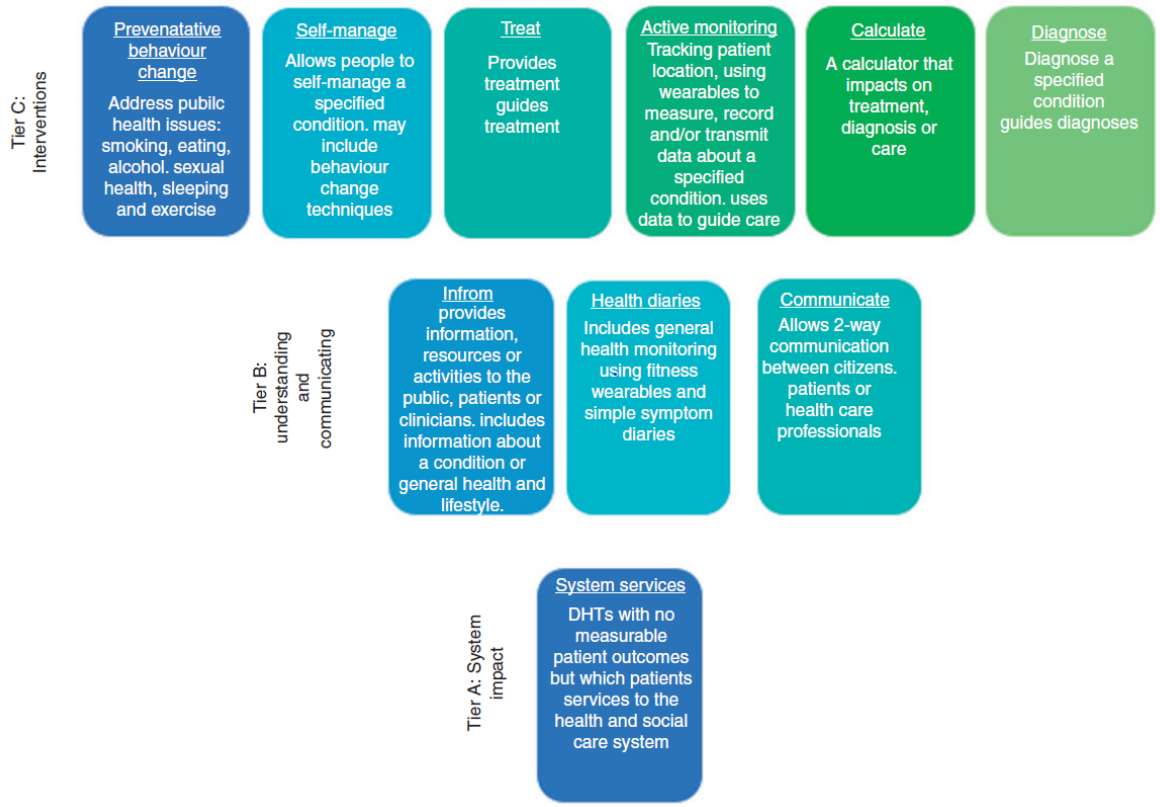


Figure 1. Ten functional categories describe the functions of the DHTs most frequently Commissioned by the health and care system (Unsworth *et al.*, 2021).

II.1.2.2 WHO's classification of digital health interventions

A variety of digital and mobile technologies are being employed to meet the demands of the health care system, and these applications may be broken down into several categories by consulting the taxonomy of digital health interventions (DHIs) (Alexandridis *et al.*, 2016) which was Initiated by the World Health Organization. This classification system, also referred to as a taxonomy, is based on the concept of a "digital health intervention," which is defined as the specific use of digital technology with the goal of improving health outcomes.

II.1.2.3 Academic research on the classification challenges of digital health interventions

The advent of eHealth held up the possibility of giving patients more power in their care, which would need more collaborative decision making (Eysenbach, 2001). One word inside the eHealth umbrella is telemedicine, which is already in use in many countries (Vyas and

Bhargava, 2021) to do things like monitor patients at home or link up medical experts and patients online to increase accessibility and quality of treatment (Sood *et al.*, 2007). Applications for mobile telemedicine enable for the customization of preventative messages and interventions (Holmen *et al.*, 2017), making them well suited for behaviour change initiatives (DiFilippo *et al.*, 2015). However, a shared and clear understanding of words in the field is necessary for the introduction of digital applications in preventative behaviour modification and integrated chronic care.

II.1.2.4 Ontology for telemedicine

To further understand how telemedicine might be classified, a study was conducted in 2020 (Otto *et al.*, 2020) to extract relevant ideas using ontologies. The provided ontology contains keywords that explain the care delivery process and the geographical context of various applications related to health technology. Figure 2 represents ontology for telemedicine and related terms. According to this study, numerous related terms were identified such as telehealth, mhealth, ehealth, digital health (and care), health IT, ICT, digitization, care models, integrated care, health smart home, smart home, and ambient assisted living, as can be seen in the figure.

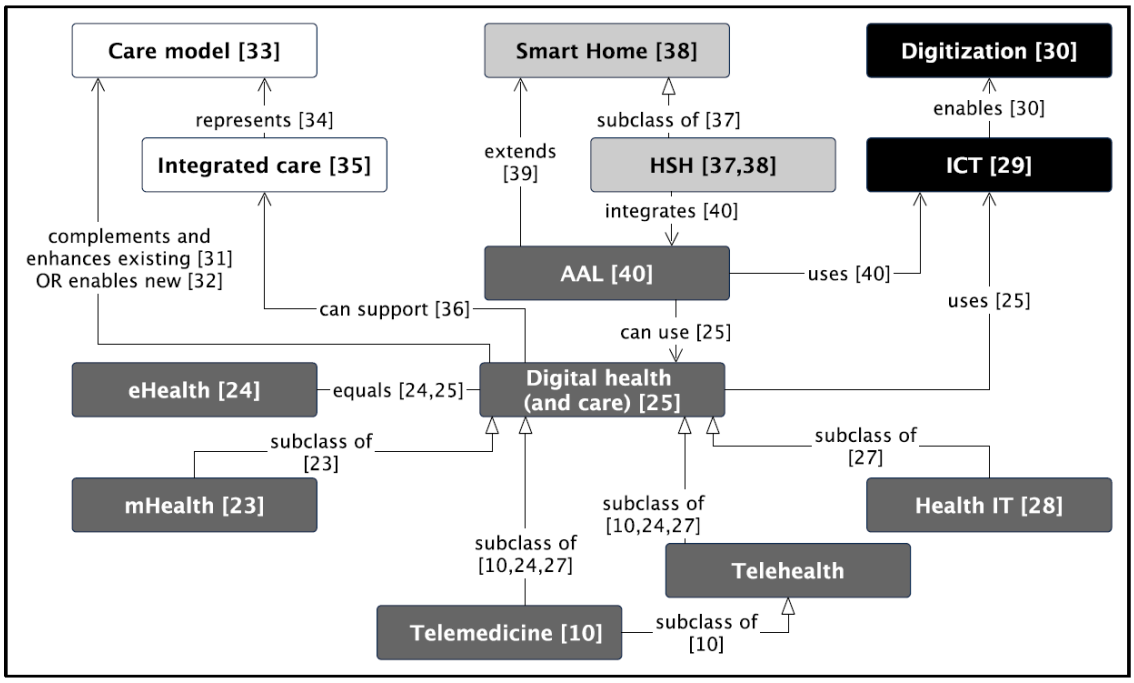


Figure 2. Ontology for telemedicine and related terms (Otto *et al.*, 2020)

II.1.2.5 Using the NICE Evidence Standards Framework for mobile health applications

NICE created a methodology in 2018 to evaluate the degree of evidence for digital health technologies (DHTs) based on their clinical function.

According to the results of one study (Nwe *et al.*, 2020) the existing version of the NICE evidence standards framework for DHTs did not allow mHealth researchers to consistently and unambiguously identify the mobile digital health apps included in the NHS app library according to their functional level.

II.1.2.6 Classification of digital health interventions from health system research perspective

A study was conducted to characterize the ideas that inform and explain eHealth deployment according to the typology developed by Sovacool and Hess, 2017 for theories of sociotechnical transformation. The results of this study showed that there are 36 different causes for eHealth deployment tactics. The two models that were used the most often were the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT).

II.1.3 WHO's general classification systems for health conditions, outcomes and interventions

In addition to the proposed methodologies, frameworks, and research to categorize clinical research explained above, three tools, namely ICD, ICF, and ICHI, have recently been introduced by the World Health Organization (WHO).

II.1.3.1 International Classification of Diseases (ICD)

Since the human condition is hard to standardize and quantify, the International Classification of Diseases (ICD) was developed as a tool for categorizing patient data. When compared to its predecessor, ICD-10, ICD-11 is a major advance. There are 55,000 separate codes for injuries, illnesses, and reasons of death (The Lancet, 2019).

II.1.3.2 International Classification of Functioning, Disability and Health (ICF)

The International Classification of Functioning, Disability and Health (ICF) is also widely recognized as the standard terminology and framework for defining human functioning on a worldwide scale (WHO, 2017b).

II.1.3.3 International Classification of Health Interventions (ICHI)

The World Health Organization (WHO) has chosen to incorporate all health treatments in a new categorization known as the International Classification of Health Interventions (ICHI) (Paviot *et al.*, 2011; Aljunid *et al.*, 2015). ICHI covers diagnostics, medicine, surgery, mental

health, primary care, medical aids, functional support, rehabilitation, traditional medicine, and public health (Fung *et al.*, 2021).

II.1.3.4 Using the ICD system academic research

Numerous studies have used these tools to categorize populations, interventions, or outcomes, most of which fall outside the scope of digital health. For example, some studies used only the ICD. For instance, one systematic review in 2017 in alcohol use (Rehm *et al.*, 2017). Another systematic review in mental and neuropsychiatric manifestations of SARS, MERS, and COVID-19 (Rogers *et al.*, 2020). Another systematic review utilized ICD-10 to identify illness regions. The purpose of this research was to assess the effect of SARS-CoV-2 on the mental health of hospital-based healthcare personnel. Six categories were identified, including depression, anxiety, acute stress reaction, post-traumatic stress syndrome, insomnia, and occupational burnout (Sanghera *et al.*, 2020).

II.1.3.5 Using the ICF system in academic research

ICF tool also was used in some systematic reviews in physiotherapy treatments on balance in multiple sclerosis patients (Paltamaa *et al.*, 2012), parameters related with adult wheelchair user engagement in social and community activities (Manuscript, 2017), and the impact of functional electrical stimulation of ankle dorsiflexors on walking in children and adolescents with spastic cerebral palsy (Moll *et al.*, 2017) A digital health related systematic review was conducted in 2021 (Bonnechère *et al.*, 2021) to compile a list and description of the various mHealth options for multiple sclerosis self-assessment and rehabilitation, as well as to define the level of evidence supporting these interventions for functioning problems classified within the International Classification of Functioning, Disability, and Health (ICF). Using the ICF technique, a variety of outcomes were identified: 53% of studies provided cognitive outcomes, 37% fatigue outcomes, 33% quality of life outcomes, 23% motor function outcomes, and 20% activity level outcomes.

II.1.3.6 Using ICHI in academic research

The ICHI instrument has also been the subject of some other studies in recent years. For instance, one study sought to review the available evidence on treatments that improve the satisfaction of persons with cognitive impairments and their caregivers with health care services (Fänge *et al.*, 2020). Using the ICHI methodology, numerous interventions were discovered. ICF and ICD were used only in one study. In this guideline (Arundale *et al.*, 2018), ICD-10 and ICF methods were used to classify illnesses and outcomes in a clinical practice guideline for the prevention of knee and anterior cruciate ligament injuries via exercise.

II.1.4 Conclusion

Although WHO tools such as ICD-11, ICHI, and ICF have predominantly been applied in non-digital health domains and are recognized for their comprehensiveness, a research gap exists concerning their adequacy in addressing the efficacy in digital health technologies, particularly digital biomarkers. The extent to which these tools can encompass the diverse spectrum of digital health technologies, including digital biomarkers, remains unknown.

II.2 Methodological quality of systematic reviews

II.2.1 The importance of methodological quality in evidence synthesis

Keeping up with primary research has become almost difficult for healthcare practitioners and policymakers due to the exponential growth of biomedical publication (Bastian, Glasziou and Chalmers, 2010a). This is why systematic reviews are such an important resource for those making health care decisions (Mulrow, 1994). Systematic reviews provide the opportunity to base judgments on accurate, precise, reliable, and exhaustive summaries of the best available information on a given topic(Mulrow, 1994). Although the Cochrane Collaboration Handbook is an excellent resource for those producing reviews, it lacks a streamlined critical appraisal tool for completed reviews (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022b). Several instruments have been created to evaluate individual research for inclusion in systematic reviews or to indicate how certain processes should be carried out (Whiting *et al.*, 2003; Wong, Cheung and Hart, 2008). Nevertheless, there are a paucity of tools that evaluate all key review techniques (Whiting *et al.*, 2016).

II.2.2 AMSTAR-2: A Measurement Tool to Assess Systematic Reviews

AMSTAR (A Measurement Tool to Assess Systematic Reviews), which was released in 2007, is one of the most extensively used instruments (Shea *et al.*, 2009). AMSTAR-2 consists of 16 questions with the answers "yes" or "no" or, for some questions, with the answer "partially yes."

II.2.3 Quality assessment of systematic reviews in digital health using AMSTAR-2

Several studies have assessed the methodological quality of systematic reviews in digital health research. Studies on methodological quality in digital health research include Dumit *et al.*, 2018 on eHealth technologies' impact on vaccination practices, with one low and five moderate quality reviews. Li *et al.*, 2020 assessed telemonitoring therapies for COPD, finding two high, two low, and four critically low quality reviews. Finucane *et al.*, 2021 employed meta-review methodology, identifying mostly moderate quality SRs on digital health interventions for palliative care. Studies on e-health and m-health interventions for weight-related behaviors in

children and adolescents (Jamshidi, Heyvaert and Van den Noortgate, 2017), digital interventions for behavioural and health outcomes in non-clinical adult groups (Gold *et al.*, 2021), synchronous implementation of digital mental health (Villarreal-Zegarra *et al.*, 2022), and efficacy of envelopes on Cardiac Implantable Electronic Devices CIED infection prevention (Motahari-nezhad, Miribonjar and Sadeghdaghighi, 2022) all noting critically low methodological quality in most reviews. An article assessed the attitudes, safety, and implementation of telehealth services in surgery, as well as the use of telehealth in Australia between 2020 and 2021. Seventeen SRs were included in this study. This study does not report the overall methodological quality of the included systematic reviews (SRs); rather, a table presents the findings of each AMSTAR-2 questions (Smith *et al.*, 2021).

II.2.4 Experience with AMSTAR-2 outside the digital health domain

Reviews on traumatic dental surgery (Magno *et al.*, 2020) and prevention approaches for early childhood caries (Soares *et al.*, 2021) mostly had low and critically low methodological quality, respectively. Methodological quality assessment of reviews on the efficacy of physiotherapy treatments for the management of tendinopathy as a single clinical entity (Girgis and Duarte, 2020), showed that most studies were of low methodological quality. Analysis of a study evaluating the methodological quality of orthodontically induced inflammatory root resorption found that most of these studies (67.9%) were classified as of moderate methodological quality (Yassir, McIntyre and Bearn, 2021).

Among the studies that do not address digital health and do not reflect the overall methodological quality of the systematic reviews are examples such as septic arthritis and osteomyelitis (Gigante *et al.*, 2019), and reducing burnout among physicians and nurses. These studies did not reflect the overall methodological quality of the included systematic reviews, but a table presents the results of each AMSTAR-2 component.

II.2.5 Conclusion

The existing literature indicates that a majority of systematic reviews in the domains of digital health and pharmaceuticals exhibit suboptimal methodological quality, revealing a significant gap in this area. Furthermore, the absence of a meta-assessment study that specifically examines and compares the methodological quality of systematic reviews between these two domains further shows the existence of a research gap. This thesis aims to address this research gap by conducting a study to fill this gap and provide insights into the methodological quality of systematic reviews in digital health and pharmaceuticals. By doing so, this research seeks to contribute to the advancement of evidence synthesis and inform decision-making processes in both fields.

II.3 Quality of evidence

II.3.1 The importance of high quality evidence in decision-making

Evidence-based practice is an essential part of today's practice. It is so important that it is impossible to imagine contemporary health care if evidence and its quality are neglected (Szajewska, 2018a). About 62% of the papers used to develop primary care guidelines and recommendations had questionable research methods and unknown relevance to patients, according to the National Institute for Clinical Excellence (Steel *et al.*, 2014). It is therefore imperative that every effort be made to enhance the quality of evidence, for which good quality research is essential. GRADE The Grading of Recommendations Assessment, Development, and Evaluation Working Group defines quality of evidence (QOE) as confidence that the reported impact estimates are accurate and competent enough to support a specific recommendation (Guyatt, 2008). Apart from the quality of evidence, there are some standards for reporting different elements in systematic reviews. The PRISMA statement is a minimal set of elements for reporting in systematic reviews and meta-analyses that are supported by evidence. Although PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was developed with the publication of reviews evaluating the effectiveness of treatments in mind, it can also serve as the basis for conducting systematic reviews that have objectives other than evaluating the effectiveness of interventions (e.g., evaluating aetiology, prevalence, diagnosis, or prognosis) (Page *et al.*, 2021).

II.3.2 The GRADE framework

In recent years, the GRADE approach has become one of the most popular methods for evaluating the quality of evidence and providing guidance for systematic reviews. GRADE Assesses the quality of evidence of the results reported in meta-analyses in four levels: high,

moderate, low, and very low. GRADE approach has been used in a variety of studies, both digital and non-digital health research, presented in this section.

II.3.2.1 Evidence quality in digital health research

Regarding digital health studies, numerous systematic reviews and meta-analyses have assessed their own reported meta-analyses in terms of quality of evidence. A meta-analysis published in 2018 summarized the data addressing the efficacy of commercially available wearable activity monitors and smartphone apps for raising physical activity levels in stroke patients. The stated outcomes for step count in the community and inpatient rehabilitation settings were very low evidence quality, while the outcomes for physical activity were of low certainty (Lynch *et al.*, 2019). A meta-analysis assessed the effectiveness of e-Health based self-management programs for people with chronic low back pain Moderate quality data suggested that e-Health based self-management programs had a clinically meaningful impact on improving disability at immediate follow-up, while low quality evidence suggested that there was no significant difference at short-term follow-ups, although with a positive trend (Du *et al.*, 2020). Whereas studies on the risk of infection, device-related mortality, and malfunction in patients with reprogrammed pacemakers and implantable cardiac defibrillators (ICDs) (Psaltikidis, Costa and Graziano, 2021) and on the effect of using sensors to prevent falls in hospitalized patients and the elderly (Cortés *et al.*, 2021) provided moderate-quality evidence, meta-analyses on the effects of digital shared decision-making (SDM) measures on patient outcomes (Vitger *et al.*, 2021) and on the effects of e-health treatments on disease activity, self-efficacy, pain, and quality of life in patients with rheumatoid arthritis (L. Zhou *et al.*, 2022) had low and critically low methodological quality.

II.3.2.2 Evidence quality outside the digital health domain

Despite the meta-analyses on digital health that assessed their own reported outcomes in terms of quality of evidence, there are other reviews of systematic reviews that assessed the quality of evidence of the outcomes of other studies. The quality of evidence synthesis of meta-analyses evaluating the effect of envelopes to reduce CIED infections was evaluated and results showed that 60% and 40% of the outcomes had low and moderate quality of evidence synthesis, respectively (Motahari-nezhad, Miribonjar and Sadeghdaghighi, 2022). According to other literature, paracetamol (Abdel Shaheed *et al.*, 2021) and chocolate intake (Veronese *et al.*, 2019) created high and low quality of evidence, respectively.

A very low to moderate quality of evidence was also assessed in two other reviews of systematic reviews on the revision of the European Association for the Study of Diabetes

recommendations (Chiavaroli *et al.*, 2019) and on traditional Chinese nursing treatments published in the Chinese journal (Jin *et al.*, 2016).

II.3.3 Conclusion

The existing literature consistently reveals a prevalent lack of high-quality evidence across various research areas. However, a notable research gap exists as no study has yet undertaken a systematic comparison to identify potential differences between these distinct research areas. Consequently, there is a compelling need for a comparative analysis to determine whether one domain demonstrates superior or inferior evidence quality compared to the other. By conducting such an analysis, this study aims to explore the relative strengths and weaknesses in evidence generation within these research domains, ultimately contributing to the advancement of evidence-based practices and decision-making processes.

II.4 Statistical power of studies

II.4.1 Insufficient statistical power is a general concern in medical research

Statistical power is an essential aspect of research that is most helpful in the design and planning stages of the study, but must be evaluated before conclusions are drawn about the results of the study. The ability to safely reject a flawed null hypothesis is what we mean when we talk about power. Many studies are not conclusive enough (Bezeau and Graves, 2001) so their results should be called inconclusive (Keen, Pile and Hill, 2005). Power measures how likely it is that a study will lead to the correct conclusions or detect an effect if one exists (Gaskin and Happell, 2014).

Problems with study power, methodological errors, and selective reporting all plague clinical research (Munafò *et al.*, 2017). Numerous studies have insufficient sample sizes to identify clinically significant improvements, and many have additional important methodological and statistical faults (Head *et al.*, 2015). Concerns about insufficient sample size and lack of statistical power have received much attention in both primary studies (Higgins, Whitehead and Simmonds, 2011). Evidence synthesis methods, such as meta-analysis, are commonly used to guide policy decisions, clinical practice, and evidence-based medicine since they are considered to provide the strongest available evidence (Gopalakrishnan and Ganeshkumar, 2013). Given that systematic reviews and meta-analyses serve as primary sources for clinical decision-making and increase the power (Cohn and Becker, 2003), it is important to consider the potential impact of low statistical power on these analyses. Therefore, exploring the statistical power of meta-analyses becomes an intriguing avenue to pursue.

II.4.1.1 Assessing statistical power in digital health research

Regarding digital health research, few studies have assessed the statistical power of studies. For example, the statistical power of effect sizes was assessed in two randomized controlled trials (RCTs). The aim of one study was to examine the effects of a tablet-based cognitive rehabilitation program on cognitive function, cognitive complaints, fatigue, and psychological distress in patients with primary brain tumors. A power analysis was conducted to determine the minimum sample size required. The analysis revealed that a group size of 50 subjects with a total of 100 participants was required, which limited the statistical power of this study (van der Linden *et al.,* 2021). A randomized controlled trial (RCT) also tested an Internet-based, parent-cantered healthy lifestyle program for preschool-aged children with a body mass index (BMI) of at least the fifty-first percentile of their age and gender. Based on their preliminary research, they hypothesized that the BMI effect size of the study would be approximately 0.4. With a predicted failure rate of 15%, 160 subjects (80 per group) should be enrolled in the study (Hammersley *et al.*, 2019).

II.4.1.2 Assessing statistical power outside the digital domain

As for non-digital health, more studies have been conducted to assess statistical power. Thirty studies assessing the cognitive effects of deep brain stimulation of the subthalamic nucleus in Parkinson's disease were reviewed, and their statistical power was calculated in a study published in 2006. After surgery, patients' ability to communicate effectively was studied, and the average effect size was 0.23, yielding an observed power of 0.16 (Woods *et al.*, 2006). Evaluate perioperative factors and unfavourable outcomes for patients receiving laparoscopic hepatectomy vs open approach was the purpose of another review published in 2011. The power analysis showed that only one research demonstrated statistical power of 80% for all four outcomes, whilst four studies lacked adequate statistical power (Mizuguchi *et al.*, 2011). In a research, the power of meta-analyses of anesthesiologic therapies was evaluated. This research indicated that just 12% of meta-analyses have a power of 80% (Imberger *et al.*, 2015). A systematic review investigated the significance of Bowel Preparation Regimen in Patients Scheduled for Colonoscopy. The findings indicated that the statistical power of qualifying included studies varied from 0.1% to 99.9%. The majority of outcomes in all qualifying trials had power below 50% (Song *et al.*, 2016).

II 4.2 Conclusion

Sample size estimation and power analysis are crucial considerations in clinical studies, yet their comprehensive reporting remains inadequate. Particularly in the realm of digital health

research, such as digital biomarkers, evaluating study power is essential from a regulatory standpoint. Compliance with regulatory requirements necessitates a thorough assessment of study power to ensure robustness and reliability of findings.

II.5 Publication bias in meta-analyses

II.5.1 Publication bias is a general threat to the quality of evidence syntheses

Research with more conclusive results is more likely to be published, leading to publication bias in meta-analyses of previously published studies. However, the likelihood of a study being published in a journal is usually related to the statistical significance of its results (Sutton and Higgins, 2008). Because of the potential for publication bias to influence the results of systematic reviews, identifying this issue is crucial (Sutton and Higgins, 2008).

II.5.1.2 Assessing publication bias in digital health

Some previously published meta-analyses in digital health have assessed the impact of publication bias in reported effect estimates using the trim-and-fill method. While some research reported only the change in effect estimate size, other studies also reported the number of missing studies in their meta-analyses using the trim-and-fill method. For example, one study examined dropout rates from randomized controlled trials (RCTs) of smartphone applications for treating depressive symptoms. The overall percentage of dropouts was 26.2%. This percentage increased to 47.8% when publication bias was taken into account (Sutton and Higgins, 2008). The purpose of another meta-analysis was to evaluate the effectiveness of mobile-health (mHealth) treatments in increasing glycemic stability and quality of life in people with type 1 diabetes (T1D). Concerning publication bias in research that investigated glycemic stability, the trim and fill approach revealed that the corrected effect size was 0.41 by including two missing studies. Regarding publication bias in studies measuring life satisfaction, the trim-and-fill technique revealed that the corrected impact estimate was 0.25 following the addition of two missing studies. Regarding publication bias, according to the trim-and-fill test, no missing studies should be added to the studies that investigated diabetic anxiety (Sutton and Higgins, 2008). The use of cell phone applications to support lifestyle modification in different diabetes subtypes was assessed in a meta-analysis. In spite of employing the trim and fill strategy to correct for publication bias after imputing four missing studies, the impact size remained statistically significant (Sutton and Higgins, 2008).

II.5.1.3 Assessing publication bias outside the digital health domain

One study aimed to define aromatherapy and sleep quality and to uncover the quantifiable effects of aromatherapy on adult sleep. Using the trim-and-fill technique, the effect size was

adjusted from 0.60 to 0.26 (Her and Cho, 2021). In some studies, only the adjusted effect size was reported. One meta-analysis examined the association between social support and mental health problems during pregnancy. The pooled odds ratio for low social support was 1.18 (95% CI: 1.01, 1.41) when the random-effects model was used to adjust for publication bias (Bedaso *et al.*, 2021). On the other hand, some other meta-analyses simply stated that no significant difference in the magnitude of their effect sizes was found with the trim-and-fill method. Studies on the effects of music-based movement therapy on Parkinson's disease patients (Zhou *et al.*, 2021) and effects of exercise in older persons (Schuch *et al.*, 2016) concluded that no major differences were discovered using the trim and fill approach in the sizes of effect estimates.

On the other hand, some other non-digital health meta-analyses have reported the number of missing studies and the change in the size of their effects (adjusted effect size). While a meta-analysis comparing blood vitamin D levels in patients with systemic lupus erythematosus and healthy controls using the trim-and-fill method found no missing (imputed) studies and no changes in effect size (Islam *et al.*, 2019), some other meta-analyses found. For example, meta-analyses on the dose-response association between clinical frailty scale and mortality in patients with COVID -19 (Pranata *et al.*, 2021) and thyroid autoimmunity during pregnancy and after delivery and on the association with postpartum depression (Minaldi *et al.*, 2020) found two missing studies whose effect estimates minimally declined.

II.5.2 Conclusions

The assessment of publication bias is a customary component of the systematic literature review (SLR) process, and it continues to be a significant concern within both the digital health and non-digital health domains. However, there exists a research gap in the comparative analysis of publication bias between these two areas, leaving the question unanswered as to which domain is more susceptible to its influence. Therefore, it is imperative to conduct an investigation to discern and compare the extent of publication bias in these distinct research domains, ultimately enhancing our understanding of potential biases and their implications for evidence synthesis and decision-making.

III. RESEARCH HYPOTHESES

The hypotheses of this thesis are listed below. Each hypothesis was tested in the chapters:

III.1 Hypothesis I.

Among systematic reviews of digital biomarkers identified in 2019-202 in the PubMed and the Cochrane library databases, in at least 95% of cases the population, intervention, outcomes, and digital biomarkers could be classified using ICD-11, ICHI, and ICF tools.

The detailed information regarding this hypothesis is provided in chapter IV. To test H1, we calculated the 90% confidence interval for the proportion of "Yes" options. If the upper limit of the 90CI < 0.95, we reject H1 with one-tailed $p<0.05$. Rejection of H1 suggests that the WHO classification systems fail to full categorize digital biomarkers.

III.2 Hypothesis II.

The statistical power of direct DBMs-based studies is significantly higher than that of indirect DBMs.

The detailed information about this hypothesis can be found in chapter (V). First, the statistical power of studies of the two groups were calculated using power analyses. Then, the normal distribution of the data was assessed using the Shapiro Wilk test. The independent sample t-test or Mann Whitney U test were applied to test the hypothesis considering the normality or non-normality of the data.

III.3 Hypothesis III.

The methodological quality of DBM-based systematic reviews in a matched sample identified by PubMed is significantly lower than that of non-DBM or pharmaceuticals.

The detailed information of this hypothesis can be found in chapter VI. First, the methodological quality of systematic reviews of the two groups were determined using the AMSTAR-2 tool, which are reported as quantitative variable between 1 and 16. Then, the normal distribution of the data was assessed using the Shapiro Wilk test. Considering the normality or non-normality of the data, the independent sample t-test or the Mann Whitney U test was applied to test the hypothesis.

III.4 Hypothesis IV.

We hypothesized that in a matched sample of systematic reviews identified in the PubMed database, the proportion of meta-analyses of DBM-based interventions with low or very low quality evidence is higher than that of non-DBMs or drug-based interventions.

The detailed information regarding this hypothesis is provided in chapter VII. First, the quality of evidence from the meta-analyses of the two groups was assessed using the GRADE instrument. If the quality of evidence of a meta-analysis is low or very low, a value of "yes" was assigned; otherwise, a value of "no" is assigned for the meta-analysis. The Fisher exact test was then applied to test the hypothesis.

III.5 Hypothesis V.

We hypothesized that in a matched sample of systematic reviews identified in the PubMed database, when using the trim and fill test for publication bias, the proportion of meta-analyses of digital biomarker-based interventions missing studies is higher than that of non-DBMS or drug-based interventions. The detailed information about this hypothesis has been provided in chapter VIII. First the number of missing studies in meta-analyses of the two groups were calculated using the trim and fill method. Then, the normal distribution of the data were assessed using the Shapiro Wilk test. Considering the type of normal distribution, the independent sample t-test or the Mann Whitney U test was applied to test the hypothesis.

III.6 Hypothesis VI.

We hypothesized that in a matched sample of systematic reviews identified in the PubMed database, when using the trim and fill test for publication bias, the difference between the calculated effect size and the adjusted effect size in meta-analyses of DBMs is significantly higher than that of non-digital biomarkers.

The detailed information concerning this hypothesis can be found in chapter VIII. First the adjusted effect size of each meta-analysis in the two groups was calculated using the trim and fill method. Then, the change between the reported effect size in meta-analyses and adjusted effect size were measured. Then, the normal distribution of the data was assessed using the Shapiro Wilk test. Considering the normal or non-normal distribution of the data, the independent sample t-test or the Mann Whitney U test was applied to test the hypothesis.

IV. ESTABLISHING A CLASSIFICATION SYSTEM IN DIGITAL BIOMARKERS

IV.1 Introduction

In the practice of evidence-based medicine (EBM), it is generally recognized that the formulation of the research question is the most significant and vital aspect of research integrity (Eldawlatly *et al.*, 2018). Most academics adhere to the PICO (population, intervention, comparison, and outcome) paradigm when formulating research questions and conducting literature reviews (Schardt *et al.*, 2007; Farrugia *et al.*, 2010). This approach allows researchers to better integrate the many components of their research topic into a coherent whole (Snowball, Library and Hospital, 1997). It is common knowledge that the creation of a well-focused research topic that includes well-articulated PICO components is the first step in designing a high-quality study (Oxman, 1993).

As suggested in the literature review, the World Health Organization family of international classifications (WHO) is an integrated set of classifications that serve as a global language for health information and consists of three reference classifications: the International Classification of Diseases (ICD), the International Classification of Functioning, Disability and Health (ICF), and the International Classification of Health Interventions (ICHI) (Fung *et al.*, 2021). The use of ICD to classify illnesses has expanded dramatically in recent years (Harrison *et al.*, 2021). ICF is widely recognized as the standard terminology and framework for defining human functioning on a worldwide scale (WHO, 2017a). For the purposes of statistics, quality, and payment, ICHI is also a standardized method for reporting and assessing health interventions among governments, service providers, managers, and researchers (WHO, 2020b). Therefore, we conducted a review to assess the usability of ICD-11 (the most recent version of ICD) for categorizing disease domains, ICHI for categorizing interventions, and ICF for categorizing outcomes in systematic reviews of digital biomarker-based studies. As explained earlier, digital biomarkers are behavioral or physiological data. Therefore, we assessed the usability of the ICF tool for categorizing behavioral and physiological data in systematic reviews of digital biomarker-based studies, as this is a tool for the classification of functioning, disability and Health.

As shown in the literature, these tools are widely used in pharmacological studies and can be used to categorize populations, interventions, and outcomes. Therefore, this study hypothesized that these tools are capable of categorizing populations, interventions, outcomes, and behavioural or physiological data of systematic reviews of digital biomarker-based studies. IF the tools can be used in formulating the research questions in the style of PICO and categorizing behavioural/physiological data in the field of digital biomarker research, this will lead to integration of digital biomarker research and improve the quality of systematic reviews and meta-analyses in the field.

IV.2 Methods

This section is based on a research paper published at the JMIR mHealth and uHealth journal (Motahari-Nezhad *et al.*, 2021). Digital biomarkers are behavioral/physiological data acquired by digital devices such as smartwatches (Nam *et al.*, 2019), heart rate, physical activity, and number of steps, according to the definition (Babrak *et al.*, 2019). In this research, digital biomarkers were defined as behavioral/physiological data and the digital devices used to assess them. Digital devices that do not objectively measure physiological or behavioural data were omitted from this investigation. For more information regarding the search strategies, inclusion and exclusion criteria, and screening, please refer to the published paper (Motahari-Nezhad *et al.*, 2021).

The individual studies that met the inclusion criteria within the identified systematic reviews (SRs) were extracted, and subsequent to the removal of duplicate records, the remaining studies were deemed eligible for final analysis. The ICD-11, ICHI, and ICF tolls were used to categorize populations, interventions, and outcomes, respectively. Digital biomarkers (behavioural/physiological data) was also categorized using the ICF tool.

IV 2.2 Statistical analysis

As explained, we hypothesized that in at least 95% of the cases populations, interventions, outcomes, and digital biomarkers can be categorized using ICD-11, ICHI, and ICF tools. For testing the hypothesis, we generate an indicator variable, which takes the value of 1 (“yes”) if

A: All populations can be categorized using the ICD-11 tool AND

B: All interventions can be categorized using the ICHI tool AND

C: All outcomes can be categorized using the ICF tool AND

E: All DBMs can be categorized using the ICF tool, otherwise 0 (No).

To test H1, we calculated the 90% confidence intervals for the proportion of "Yes" options. If the upper limit of the 90CI < 0.95, we reject H1 with one-tailed $p<0.05$. Rejection of H1 suggests that the WHO classification systems fail to full categorize digital biomarkers.

IV.3 Results

IV.3.1 Search and screening results

After conducting a thorough search on electronic databases, 31 systematic reviews were identified that met the inclusion criteria for digital biomarkers. We invite the readers to peruse our previously published article for a comprehensive exposition of the final studies that were included (Motahari-Nezhad, Fgaier, *et al.*, 2022). Refer to figure 3 for further details regarding the search and screening process of the articles.

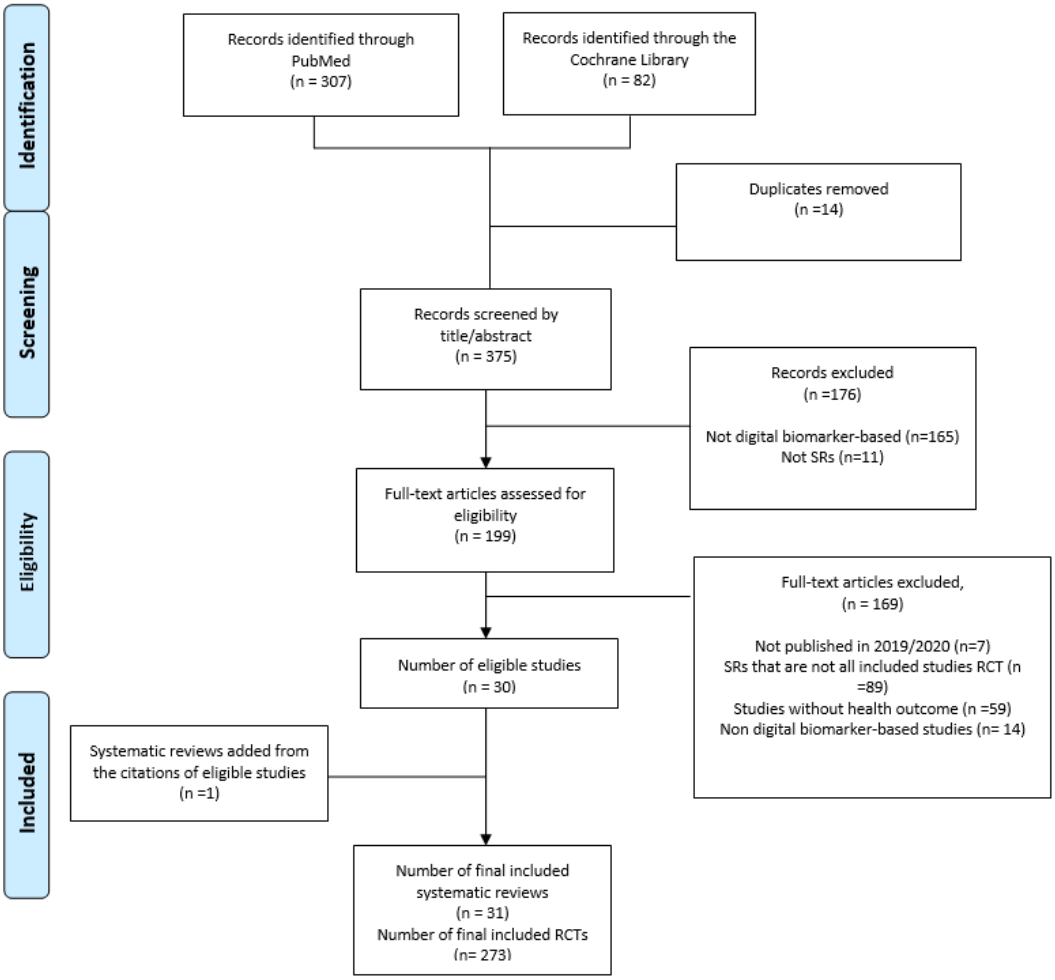


Figure 3. PRISMA diagram of the selection and screening process for a classification system in DBMs

From these systematic reviews, a total of 335 single randomized controlled trials (RCTs) were extracted. After removing duplicates, 273 RCTs remained for the final analysis.

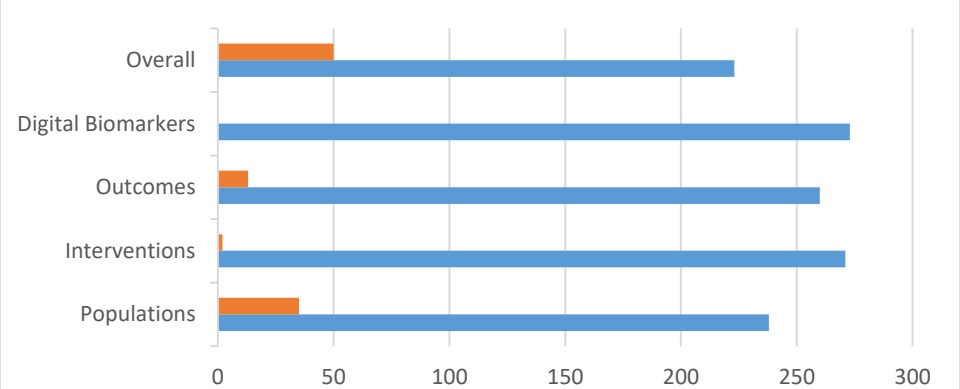
IV.3.2 Classification of studies using the WHO tools

The populations, interventions, and outcomes of these RCTs were examined to determine if they could be categorized using the World Health Organization (WHO) tools ICD-11, ICHI, and ICF, respectively. In addition, the digital biomarkers, which consist of both behavioral and physiological data, were also classified using the ICF tool.

Out of the total 273 RCTs analysed, 35 studies had populations that could not be classified using the ICD tool, as these studies included non-clinical populations or healthy individuals without any specific clinical conditions (Jakicic *et al.*, 2016; Rote, 2017). Additionally, two RCTs, their interventions including electronic health record (Ryu *et al.*, 2017) and blood cholesterol monitoring (Thorndike *et al.*, 2014; Lim *et al.*, 2016), could not be categorized using the ICHI tool.

Regarding outcomes, seven specific outcomes in thirteen RCTs, including motivation and adherence rate (Griauzde *et al.*, 2019), number of ICD shocks (Lüthje *et al.*, 2015), defibrillation energy requirements (Stevens *et al.*, 1996), number of visits (Landolina *et al.*, 2012), depression and anxiety (Blough and Loprinzi, 2018), stroke and bleeding (Martin *et al.*, 2015), and rates of hospital admissions (Abraham *et al.*, 2016), could not be coded using the ICF tool.

However, all digital biomarkers, including both behavioral and physiological data, were able to be classified and coded using the ICF tool. Figure 4 shows the applicability of the WHO tools in categorizing population, intervention, outcome and digital biomarkers of the studies.



	Populations	Interventions	Outcomes	Digital Biomarkers	Overall
No	35	2	13	0	50
Yes	238	271	260	273	223

Figure 4. The applicability of WHO tools in categorizing the populations, interventions, outcomes, and digital biomarkers of the included RCTs.

In accordance with the methodology, the confidence intervals of the variables were calculated at a 90% significance level. As presented in the tabulated data in table 1 and figure 5, the computed confidence intervals were determined to be higher than 0.95 for interventions, outcomes, and digital biomarkers. It is therefore evident from the outcomes that since the upper limits exceed 0.95, it may be reasonably inferred that the WHO tools including ICHI, and ICF are valid instruments for characterizing interventions, outcomes, and digital biomarkers separately

Based on the findings of our analysis, it is determined that the upper confidence interval for populations does not exceed the threshold of 0.95. Consequently, this leads to the conclusion that the utilization of ICD-11 for the purpose of categorizing populations in studies is not feasible. Furthermore, when accounting for factors population, intervention, outcome, and digital biomarkers together in the analysis, the upper confidence interval is computed to be 0.85, indicating a value lower than the established threshold of 0.95. As a result, it can be inferred that the tools provided by the World Health Organization lack the capacity to adequately address the formulation of populations, interventions, outcomes, and digital biomarkers together in studies related to digital biomarkers. Accordingly, based on the result we cannot accept the hypothesis that among systematic reviews of digital biomarkers identified in 2019-2020 in the PubMed and the Cochrane library databases, in at least 95% of cases the

population, and intervention, and outcomes, and digital biomarkers could be classified using ICD-11, ICHI, and ICF tools.

Table 1. Usability of ICD-11, ICHI, and ICF Tools for Categorizing Populations, Interventions, Outcomes, and Digital Biomarkers

Variable	Observations	Proportion	Standard error	90% confidence intervals
Population	273	0.87	0.02	0.84 – 0.9
Intervention	273	0.99	0.005	0.98 - 1
Outcome	273	0.95	0.013	0.93 – 0.97
Digital biomarkers	273	1	0	1 - 1
Overall	273	0.82	0.023	0.78 – 0.85

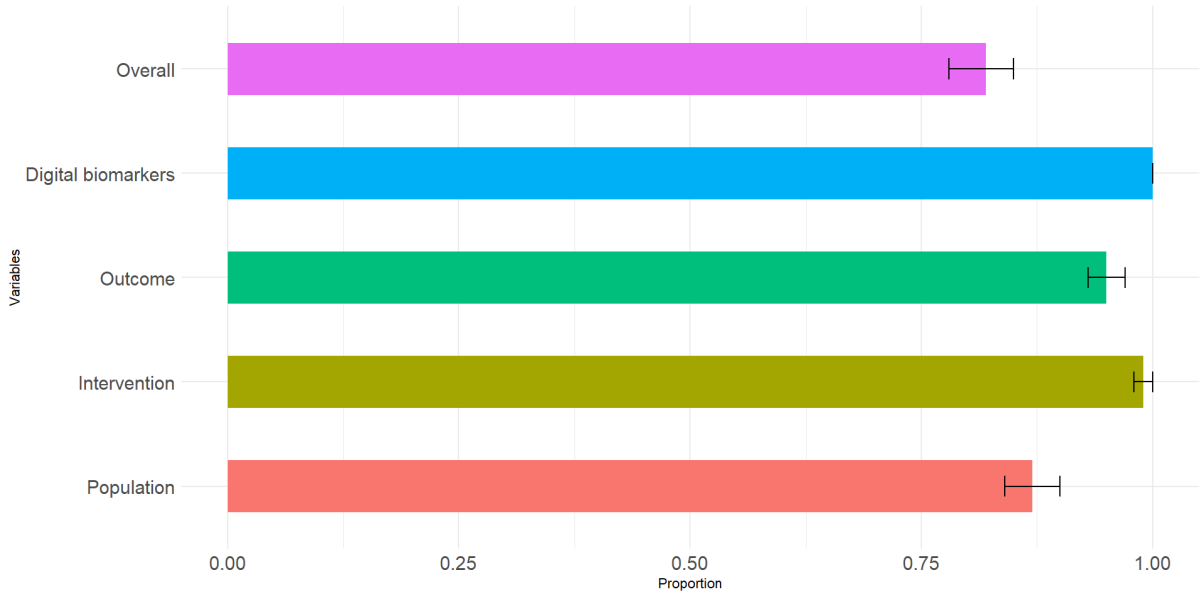


Figure 5. Proportions and 90% confidence intervals for population, intervention, outcome, and digital biomarkers in terms of WHO usability in categorizing studies

IV.3.3 Additional analysis

In light of the fact that 35 studies encompassed general populations without any disease or illnesses, we excluded such studies to ascertain the suitability of WHO tools in categorizing studies within the context of patients with specific illnesses. As a result, an upper limit was computed for the revised dataset, yielding a measurement of 0.96. Consequently, it can be deduced that the WHO tools possess the efficacy to classify populations, and interventions, and outcomes, and digital biomarkers within the subset of populations with specific illnesses, but WHO tools are not applicable in digital biomarkers including healthy people.

IV.4 Discussion

Based on the eligibility criteria, 273 RCTs met the inclusion criteria. The findings of this study demonstrated that the utilization of WHO tools for the formulation of digital biomarker studies is not applicable. Conversely, the analysis revealed that the WHO tools exhibit efficacy in categorizing the components of digital biomarker studies when conducted on populations with specific diseases (non-general populations). Given the rapid proliferation of digital biomarker health technologies (Coravos, Khozin and Mandl, 2019), which are predominantly utilized by the general public for preventative purposes (Park *et al.*, 2019), it is noteworthy that the existing WHO tool, namely ICD-11, is not applicable to effectively categorize and account for such general populations. Consequently, there is a potential for enhancing the utility of WHO tools in formulating the components of DBM studies by developing a dedicated code within ICD-11 specifically designed to address the general populations.

Otherwise, this difficulty can result in incomplete syntheses of evidence. The swift introduction of digital health technologies creates the challenges associated with defining such technologies. Consequently, there is a pressing need to revise existing definitions and establish explicit evidentiary requirements to facilitate widespread adoption of these novel and innovative tools (Izmailova, Demanuele and McCarthy, 2023). Failure to address this issue can ultimately impede the generation of comprehensive evidence by these emerging technologies.

The study found that all digital biomarkers, including behavioral and physiological data, could be classified and coded using the ICF tool. The ability to classify digital biomarkers using the ICF tool enables researchers and clinicians to document and analyse the relationship between digital biomarkers and health outcomes in a standardized and comprehensive manner, leading to the standardizations of the clinical research and can help to systematize the identification and use of different types of digital biomarkers in diagnosing, predicting and changing the health status of a wide range of populations with different clinical settings. The ubiquity of

digital biomarker technologies, including their rapid proliferation and broad adoption among various segments of the population, such as students (H-Jennings *et al.*, 2016), office workers (Parry *et al.*, 2013), and healthy adults (Croteau *et al.*, 2007), highlights the rationale for this phenomenon. Furthermore, it should be noted that two interventions, including electronic health record and blood cholesterol monitoring, could not be categorized using the ICHI tool. In terms of outcomes, the ICF tool was unable to code seven specific outcomes, highlighting notable gaps in the existing tools. Accordingly, the developers of such tools should take into account these gaps and make necessary adjustments.

These findings highlight the limitations and challenges of using standardized classification tools in clinical research and practice, especially in fast pacing clinical research field digital biomarkers (Coravos, Khozin and Mandl, 2019). As delineated in the literature review, alternative classification systems exist, such as the National Institute for Health and Care Excellence's (NICE) Evidence Standards Framework (ESF) (NICE, 2019) and the Ontology for telemedicine (Otto *et al.*, 2020). Nevertheless, the WHO tools enjoy widespread use and global recognition for health classification and research, whereas the ESF and the Ontology for telemedicine have a narrower scope and are primarily employed in specific contexts.

The adoption of a standardized terminology, such as the utilization of World Health Organization (WHO) tools, within the realm of digital biomarker research, can facilitate the application of deep learning methodologies for categorizing studies based on their Population, Intervention, Comparison, and Outcome (PICO) elements (Kang, Zou and Weng, 2019). This harmonization of language and frameworks enhances the compatibility between digital biomarker research and advanced computational techniques, enabling more efficient and accurate classification and analysis of studies.

IV.4.1 Strengths and limitations

The process of selecting systematic reviews (SRs) pertaining to digital biomarkers involved two independent reviewers, while a singular reviewer conducted a unified classification of individual studies within the selected SRs. Although the study has provided some insights into the utility of the WHO tools for categorizing digital biomarkers, there are some limitations that should be acknowledged. First, the study only focused on systematic reviews of digital biomarkers published in two years 2019 and 2020. Moreover, double coding using the WHO tools could improve the robustness.

IV.4.2 Conclusion

Digital biomarkers are predominantly employed in populations without pre-existing health conditions, and existing classification systems often lack coverage for healthy populations, particularly those at risk. Conversely, the World Health Organization (WHO) tools demonstrated effective classification of PICO (Population, Intervention, Comparison, and Outcome) statements for studies involving individuals with illnesses. In the rapidly evolving digital landscape, the establishment of a standardized classification system holds significant importance for medical decision-makers and payers.

V. COMPARING THE STATISTICAL POWER OF DIRECT AND INDIRECT DBMs

V.1 Introduction

Inadequate statistical power indicates that the study lacks the requisite power to detect significant effects (Pigott and Polanin, 2020). This deficiency can lead to a study that yields findings that do not accurately reflect the true magnitude of the impact, resulting in questionable outcomes (Nord *et al.*, 2017). Moreover, a low statistical power increases the risk of type II errors, which arise when a study erroneously accepts a false null hypothesis (Pigott and Polanin, 2020). This situation can lead to wastage of resources and missed opportunities for identifying genuine effects that could be of use. Furthermore, a low statistical power may affect the generalizability of the results by producing a sample size that does not represent the population under investigation. Hence, researchers, organizations, and healthcare service providers intending to assess the efficacy of a digital device, whether utilizing direct or indirect digital biomarkers, should consider the requisite sample size before conducting the study (Ebrahim Valojerdi, Tanha and Janani, 2017), to ensure the production of valid and robust outcomes (Martínez-Mesa *et al.*, 2016). On the other side, enrolling participants and patients beyond the optimal threshold in a study leads to a squandering of valuable time and resources (Button *et al.*, 2013).

There are two types of data that are measured using digital devices: physiological data and behavioral data, which can be considered as direct and indirect interventions, respectively. Direct interventions are interventions that directly affect the physiological characteristics of a patient population, whereas indirect interventions affect the behavioral characteristics of a population (Meerwijk *et al.*, 2016). For example, in the field of digital biomarkers, implantable cardiac defibrillators (ICDs) (Mahdi Abid *et al.*, 2022), which directly set patients' heart rhythms as physiological characteristics, are referred to as direct interventions, whereas physical activity trackers such as Fitbit (Ringeval *et al.*, 2020), which only measure physiological parameters such as heart rate and step count, are categorized as indirect interventions because they only measure behavioral human parameters. Indirect digital biomarkers introduce several layers of uncertainty, encompassing factors such as individuals' comprehension and response to the generated signal. Conversely, direct digital biomarkers circumvent the need for additional human factors. Furthermore, direct digital biomarkers

assume a higher-risk category due to interventions carried out by machines rather than healthcare professionals. Therefore, studies involving direct digital biomarkers may exhibit more meticulous planning and stringent methodologies and we hypothesised that direct digital biomarkers would possess greater statistical power compared to their indirect counterparts.

Because of the importance of DBM studies for medical decision making and the importance of assessing the power of these studies, we conducted a study to evaluate the statistical power of studies of digital biomarkers and to compare the statistical power between direct and indirect digital biomarkers to determine which type of digital biomarkers has more power and which type of digital biomarkers has too little power. Determining and comparing the power of direct and indirect digital biomarkers will inform clinical researchers and health policy makers about how statistically powerful these studies are and how direct and indirect digital biomarkers differ in terms of statistical power.

V.2 Methods

This research is based our article published in the Journal of Medical Internet Research. For further details regarding the search strategies, screening, inclusion and exclusion criteria refer to the published paper (Hossein Motahari-Nezhad, Al-Abdulkarim, et al. 2022). The individual studies included in the meta-analyses identified through systematic reviews (SRs) were extracted, and subsequent to the removal of duplicate records, were divided into two categories direct and indirect DBMs and their statistical power was calculated.

V.2.1 Power calculations

For continuous outcomes, the pooled standard deviation was calculated using Formula 1 in Appendix I when the means for the two groups were presented initially. Subsequently, employing Formula 2 in Appendix I (Cohen's d formula), the effect size was computed. Furthermore, employing Formula 3 in Appendix I and utilizing the pwr package in the R programming language, the statistical power was calculated. In cases where only Cohen's d was reported in a study, Formula 4 was employed to estimate the power. As for dichotomous outcomes, the STATA package Power formula 5 was utilized. All power calculations were performed considering a 95% confidence levels and an alpha value of 0.05.

V.2.2 Statistical analysis

To test the hypothesis, the average of statistical power of the direct and indirect DBMs was calculated and compared. In this section, we hypothesized that the statistical power of direct DBMs is significantly higher than that of indirect DBMs. To test the hypothesis, we first test

the normal distribution of the data between the two groups using the Shapiro Wilk test with 95% confidence intervals. In case of the normal distribution of the data in the two groups, a parametric two independent sample t-test was employed. Otherwise, a non-parametric Mann Whitney U test with 95% confidence intervals was used.

V.3 Results

V.3.1 Search and screening results

All in all, 25 SRs met the inclusion criteria. In the subsequent phase, the individual specific studies comprising a total of 273 DBM- based articles were extracted from the eligible SRs. Figure 6 shows the search and screening process of the articles.

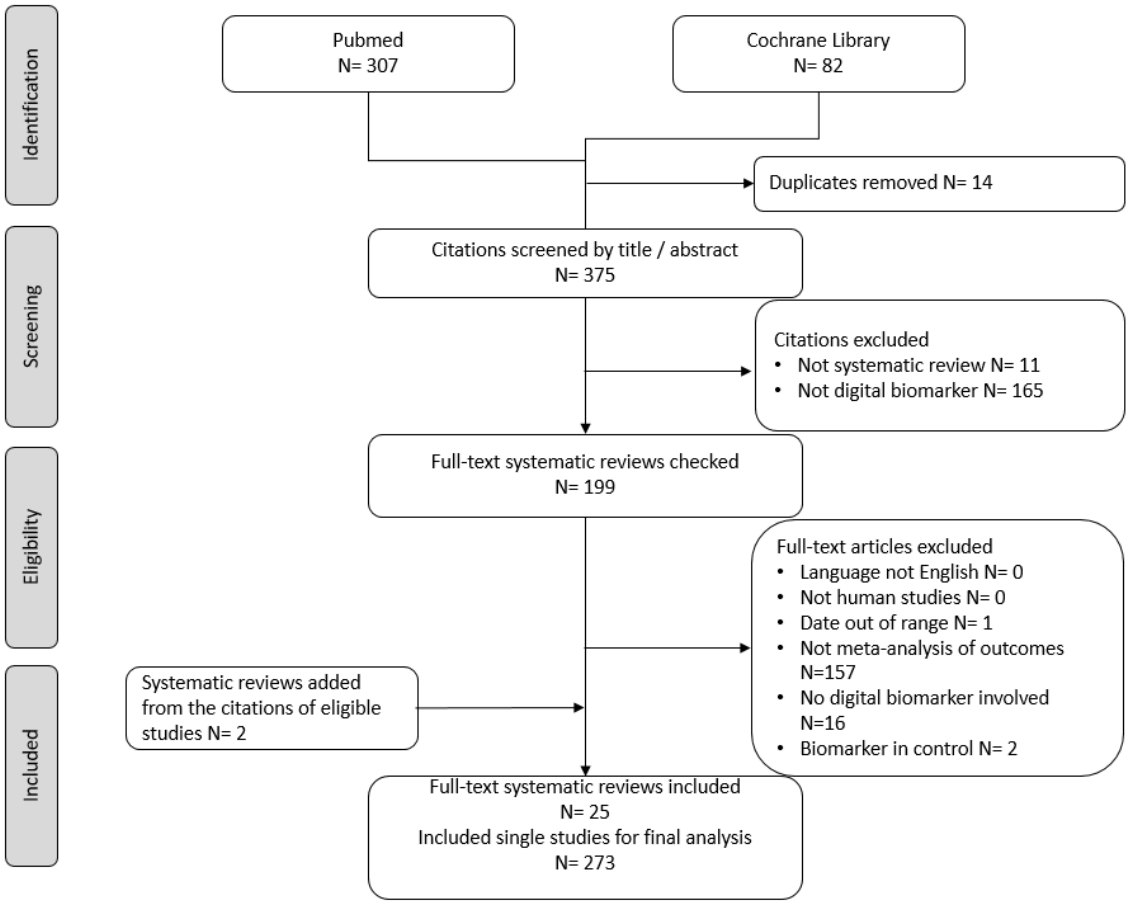


Figure 6. PRISMA diagram for selecting/screening process of direct and indirect DBMs

V.3.2 Power analysis

The identified digital biomarker based studies were categorized as direct or indirect interventions. Direct digital biomarkers-based interventions represented 16.5% or 45 of the

total studies, while indirect digital biomarkers accounted for 83.5% or 228 studies. The statistical power of each study was calculated and reported as a measure between 0 and 1.

The mean power of the direct and indirect digital biomarker-based studies were estimated as shown in table 2. For the direct digital biomarkers, the mean power was calculated to be 0.47 (SD=0.34). For the indirect digital biomarkers, the mean power was reported to be 0.44 (SD=0.34).

Table 2. Mean power of studies of direct and indirect digital biomarkers

Type of study	Frequency	Mean	Standard deviation	Ranges
Direct digital biomarkers	45	0.47	0.34	0.05 - 1
Indirect digital biomarkers	228	0.44	0.34	0.05 - 1

V.3.3 Normal distribution

The present study employed the Shapiro-Wilk W test to assess the normality of the power variable within two categories of digital biomarkers, direct and indirect. For the direct digital biomarkers, a sample of 45 observations was examined, resulting in a W value of 0.94, a corresponding z-score of 2.02, and a probability of 0.022. This outcome indicates a significant deviation from normality within the power variable for the direct digital biomarkers. Similarly, the indirect digital biomarkers were investigated using the same methodology, with 228 observations being subjected to the Shapiro-Wilk W test. The obtained W value of 0.91, accompanied by a z-score of 6.15 and a probability of 0.000, corroborated significant non-normality within the power variable for this subgroup of digital biomarkers. Therefore, the findings unequivocally indicate that the power variable does not conform to a normal distribution for either the direct or indirect digital biomarkers. See table 3 for further details. Figure 7 also visualizes the histogram graph of power of direct and indirect digital biomarkers.

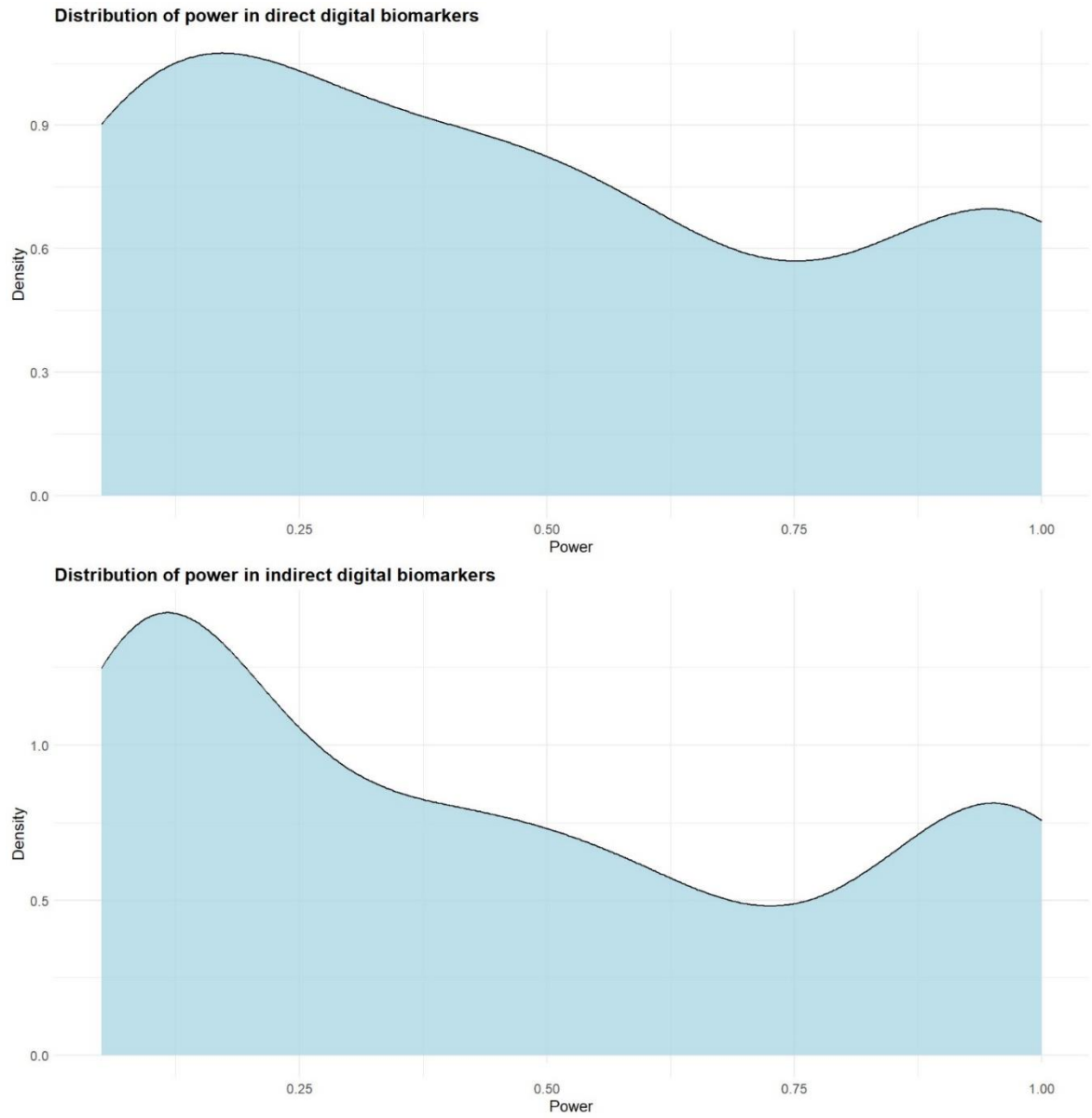


Figure 7. Histogram of power variable in studies of direct and indirect digital biomarkers

Table 3. Normal distribution testing of power variable in the studies of direct and indirect DBMs

Type of studies	Observations	W	Z-score	P-value
Direct digital biomarkers	45	0.94	2.02	0.022
Indirect digital biomarkers	228	0.91	6.15	0.000
Total	273			

V.3.4 Statistical analysis

According to the non-normal distribution of power variable, a Wilcoxon rank-sum (Mann-Whitney) test was employed to compare the power between two categories of digital biomarkers, namely direct and indirect.

The null hypothesis that the power in the direct digital biomarkers is equal to that in the indirect digital biomarkers was tested. The probability of the test was reported as 0.61, suggesting a lack of statistical significance in the difference between the two groups (table 4). Therefore, the Wilcoxon rank-sum test demonstrated no significant differences in the power between the direct and indirect categories of digital biomarkers and we reject this hypothesis at significance level 95% and conclude that there is no significant difference in statistical power between direct and indirect digital biomarker-based studies. Therefore, we reject the hypothesis that the statistical power of direct digital biomarker-based studies is significantly higher than that of indirect digital biomarkers in a sample of systematic reviews identified using PubMed and the Cochrane library in 2019-2020.

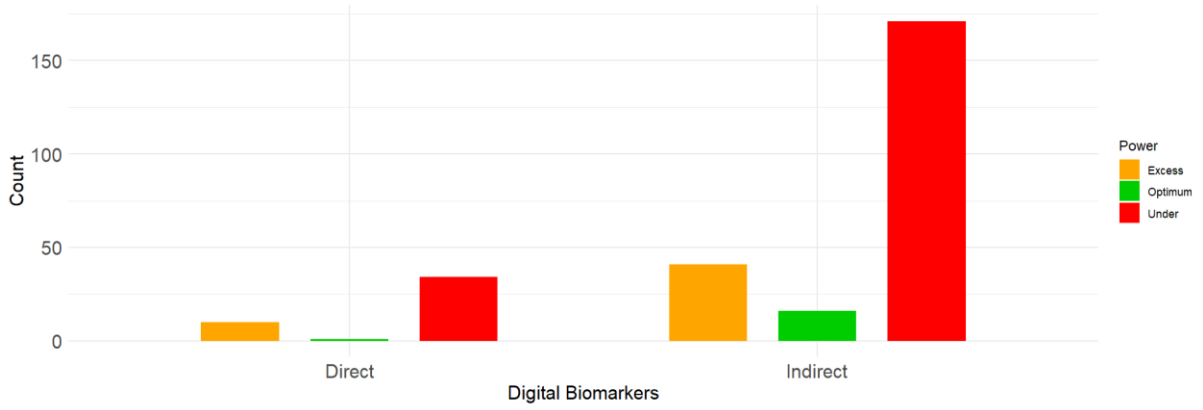
Table 4. The results of Rank sum test to determine the difference between the powers of the two groups

Type of studies	Ranks sum	Expected	Adjusted variance	Z-score	P-value
Direct digital biomarkers	6411.5	6165	233527.75	0.51	0.61
Indirect digital biomarkers	30989.5	31236			

V.3.5 Additional analyses

V.3.5.1 Grouping studies according to their power level

Given that a power level of 0.8 is conventionally considered the minimum threshold for an optimal study power (Serdar *et al.*, 2021), we classified the studies into three distinct categories. The first group, denoted as "underpowered studies," comprises those studies whose power falls below 0.8. The second group encompasses studies with power values ranging between 0.8 and 0.9, representing the range considered as the optimal power level. We refer to this category as the "optimum power" group. Lastly, the third group encompasses studies exhibiting power levels exceeding 0.9, which we designate as the "excess-powered studies" group. As observed in figure 8, the distribution of power levels in both the direct and indirect digital biomarker groups exhibits an equivalent pattern. In both groups, a substantial proportion of studies, accounting for approximately 75%, demonstrate an underpowered nature. Furthermore, a relatively small percentage of studies in both groups display an optimal power level. Additionally, the occurrence of studies with excess power is comparable, comprising 22.22% (10/45) for direct digital biomarkers and 17.98% (41/228) for indirect digital biomarkers.



Biomarkers	Under	Optimum	Excess
Direct	34	1	10
Indirect	171	16	41

Figure 8. The distribution of power types in direct and indirect DBMs

V.3.5.2 Statistical power of meta-analyses

As an additional analytical step, we also conducted an evaluation of the statistical power of meta-analyses to investigate potential distinctions between the two groups. Initially, we

extracted 95 meta-analyses from the eligible systematic reviews, which were subsequently categorized into direct and indirect groups. Subsequently, the statistical power of each meta-analysis was computed using R programming language. Refer to appendix I for further details regarding statistical power calculations of meta-analyses. The outcomes revealed that the reported statistical power of meta-analyses for direct and indirect digital biomarkers was 0.87 (SE=0.064) and 0.81 (SE=0.036), respectively. Subsequently, we examined the normality of the power distribution for both direct and indirect digital biomarker groups using the Shapiro-Wilk test. The results indicated that neither group followed a normal distribution (p-values < 0.05). Consequently, the Wilcoxon rank-sum test was employed to explore potential significant differences between the statistical power of meta-analyses for direct and indirect digital biomarkers. The findings further supported the absence of any significant difference between the statistical power of meta-analyses for direct and indirect digital biomarkers (p-value > 0.05).

V.4 Discussion

The results of this investigation indicate that there exists no statistically significant distinction in the magnitude of statistical power between studies conducted on direct and indirect digital biomarkers. Consequently, our examination did not yield evidence to support the hypothesis hypothesizing a markedly superior statistical power for studies focusing on direct digital biomarkers when compared to those on indirect digital biomarkers.

The present study demonstrated comparable statistical power between direct digital biomarker studies, mainly focused on implantable cardiac defibrillators, and indirect digital biomarker studies, primarily utilizing wearable activity trackers. The findings further revealed that the mean power values of both direct and indirect digital biomarkers were 0.47 and 0.44, respectively, indicating a relatively low power for both types of biomarkers on average. The further data analysis also indicated that the majority of studies focusing on both direct and indirect digital biomarkers were categorized as underpowered investigations. This situation can lead to wastage of resources and missed opportunities for identifying genuine effects that could be of use. Insufficient statistical power can also exacerbate the issue of publication bias, as underpowered studies are more prone to yielding non-significant results. Consequently, these studies face a higher likelihood of remaining unpublished (Nair, 2019). However, in cases where study results are statistically significant, smaller studies tend to be published, thereby introducing the potential for the small-study effect to manifest in subsequent meta-analyses.

Hence, researchers, organizations, and healthcare service providers intending to assess the efficacy of a digital device, whether utilizing direct or indirect digital biomarkers, should consider the requisite sample size before conducting the study (Ebrahim Valojerdi, Tanha and Janani, 2017), to ensure the production of valid and robust outcomes (Martínez-Mesa *et al.*, 2016). Failure to do so not only results in the inefficient allocation of time and resources but also undermines the ability to adequately establish the efficacy of the service, potentially leading to adverse implications for healthcare systems (Button *et al.*, 2013). Based on the results, a mere fraction of direct and indirect digital biomarkers demonstrated optimal power, while one-fifth of studies utilizing both types of biomarkers exhibit excessive power. Enrolling participants and patients beyond the optimal threshold in a study leads to a squandering of valuable time and resources. Hence, researchers and managers in the digital biomarker industries should adhere to an optimal power range of 0.8-0.9, while avoiding power levels exceeding 0.9, to effectively allocate time and resources (Button *et al.*, 2013).

In the context of digital biomarkers, it is therefore crucial to implement adequate sample sizes and rigorous study designs that ensure the research findings possess sufficient statistical power to detect meaningful effects. This will enhance the quality of evidence pertaining to digital biomarkers, empowering physicians to make more confident and evidence-based decisions in the best interest of their patients. In the latter section of this chapter, an examination was conducted to evaluate the statistical power of meta-analyses in light of the low power observed in both direct and indirect digital biomarker studies, alongside the absence of a significant difference between them. The findings revealed that, on average, the statistical power of meta-analyses for both direct and indirect digital biomarkers reached an optimal level. Moreover, there was no substantial disparity between the two types. Adequate meta-analyses have the potential to mitigate the issue of low statistical power inherent in individual studies. However, the attainment of reliable and robust conclusions necessitates several key factors: robust PICO (Population, Intervention, Comparison, and Outcome) classification for comparable technologies, stringent methodological quality, and crucially, the absence of publication bias within the field of investigation.

There is a research gap regarding the assessment of statistical power in meta-analyses in digital health, with limited studies conducted in this area. For example, one study aimed to investigate the effects of a tablet-based cognitive rehabilitation program on cognitive function, cognitive complaints, fatigue, and psychological distress in primary brain tumor patients. A power analysis was conducted to determine the minimum required sample size, revealing limitations

in the statistical power of the study (van der Linden *et al.*, 2021). While individual studies on direct and indirect digital biomarkers exhibited, on average, low power, the synthesis of these studies in meta-analyses effectively bolstered statistical power.

V.4.1 Strengths and limitations

The present study represents the initial attempt to evaluate the statistical power of the two primary categories of digital health intervention. To date, there has been no clear understanding of the contrast in statistical power between the two groups, and this study serves to bridge this research gap.

The present study also has some limitations that need to be considered. The study sample was restricted to systematic reviews published in PubMed and the Cochrane library in 2019-2020. Hence, the generalizability of the findings to other databases and time periods may be limited.

V.4.2 Conclusion

In light of their divergent mechanisms of action, it is recommended that researchers and healthcare industries adopt a new classification for digital biomarker technologies, distinguishing between direct and indirect types and they were regulated differently. When commencing clinical investigations to assess the efficacy of these technologies as interventions, it is imperative for researchers and healthcare industries to ensure an optimal power ranging from 0.8 to 0.9. While straying below 0.8 devastates the validity of the findings, exceeding 0.9 unnecessarily squanders time and budgetary resources. Recognizing the underpowered nature of individual clinical studies pertaining to digital biomarkers, physicians should not rely solely on such investigations. Instead, our evaluation accentuates the significance of consulting meta-analytic results, as these evidence syntheses enhance the statistical power of the study, which applied to both direct and indirect digital biomarkers.

VI. METHODOLOGICAL QUALITY OF SYSTEMATIC REVIEWS OF DIGITAL BIOMARKERS COMPARED WITH SYSTEMATIC REVIEWS OF NON-DIGITAL BIOMARKERS OR PHARMACEUTICALS

VI.1 Introduction

Optimal systematic review establishment, using best practices to reduce bias in data collection, assessment, and summarization, is essential for researchers to understand whether or not they can trust the results (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022c).

The development of systematic reviews over the past decades has raised concerns that the exponential growth in the number of published systematic reviews may have contributed to an increase in the amount of information that needs to be processed (Bastian, Glasziou and Chalmers, 2010b; Fuhr and Hellmich, 2015; Tebala, 2015; Ioannidis, 2016). Due to the rapid expansion in the field, we assumed that methodological quality of digital biomarker systematic reviews may be compromised.

According to our recently published article, most systematic reviews of digital biomarkers relate to cardiovascular diseases and physical activity (Motahari-Nezhad *et al.*, 2021). The mostly used digital biomarker for cardiac diseases was implantable cardiac defibrillators (ICDs) and mostly used digital biomarkers for physical activity was wearable activity trackers. On the other hand, there are some other interventions (non-digital biomarkers) for cardiovascular diseases and physical activity. However, the quality of systematic reviews of digital biomarkers compared with those of non-digital biomarkers is unclear. To our knowledge, there have been no published studies that systematically compared the quality of systematic reviews of digital biomarker-based interventions with that of non-digital biomarkers. Therefore, with this study, we aimed to determine the difference between the methodological quality of systematic reviews of digital biomarker-based interventions and non-digital biomarkers. In the absence of official standards and definitions for digital biomarkers, this study assumed that the methodological quality of systematic reviews of non-digital biomarkers is higher than that of digital biomarkers. The results of this research will inform researchers in the field of digital biomarkers and highlight the weaknesses and positive points of systematic reviews of digital biomarker-based interventions compared to non-digital

biomarkers in terms of methodological quality, leading to better medical decision making utilizing digital biomarkers.

VI.2 Methods

VI.2.1 Search strategy for finding ICD-based systematic reviews

To find systematic reviews of ICD-based interventions for heart failure reporting mortality outcomes, a comprehensive search of the PubMed electronic database was performed

VI.2.1.1 Screening and selection of studies

The titles/abstracts of the retrieved documents were screened according to the following criteria:

1. English language 2. Human studies 3. Systematic review.

The full texts of the remaining studies were then screened to identify the final eligible systematic reviews based on the following factors:

1. Study design: systematic review of randomized controlled trials (RCTs) with meta-analysis.

2. Population: all patients with heart failure with all other clinical conditions.

3. Intervention: implantable cardiac defibrillator 4. Comparison group: all type of comparators (non-ICDs) 5. Outcomes: Mortality in patients with heart failure.

6. Human study of any age group and gender.

VI.2.2 Search strategy for finding non-DBMS or drug-based systematic reviews

After doing full-text screening and determining the exact number of eligible studies, a similar search was conducted to find non-DBMs or drug-based interventions in heart failure patients with the same number of included studies.

To identify non-DBMs or drug therapies in patients with heart failure, another PubMed search was conducted and time of the search was unrestricted:

VI.2.2.1 Screening and selection of studies

The retrieved results were sorted by best match and their titles/abstracts and full texts were screened from the beginning one by one to find eligible studies that met the following inclusion criteria:

1. English language 2. Human studies 3. Systematic review 4. Study design: systematic review of randomized controlled trials (RCTs) with meta-analysis.

5. Population: all patients with heart failure.

6. Intervention: all non-DBMS or drugs for patients with heart failure 4. Comparison group: all kinds of comparators 5. Outcomes: Mortality in patients with heart failure.

6. Human study for any age group and gender.

VI.2.3 Search strategy for finding activity tracker-based systematic reviews

To find systematic reviews of activity tracker-based interventions, a search in PubMed was performed.

VI.2.3.1 Screening and selection of studies

The inclusion criteria:

1. English language 2. Human studies 3. Systematic review 4. Study design: systematic review of randomized controlled trials (RCTs) with meta-analysis.

5. Population: all patients with any clinical condition.

6. Intervention: wearable activity trackers 7. Comparison group: all types of comparators 5. Outcome: weight.

8. Human study of any age group and sex.

VI.2.4 Search strategy for finding non-activity tracker-based systematic reviews

After doing full-text screening and determining the exact number of eligible studies, a similar search was conducted to find non-digital biomarker-based interventions in patients with the same number of included studies. To be able to find their non-DBMs or pharmaceutical matches, weight outcomes was considered as the comparable outcome.

VI.2.4.1 Screening and selection of studies

The following factors were considered as the inclusion criteria:

1. English language 2. Human studies 3. Systematic review 4. Study design: systematic review of randomized controlled trials (RCTs) with meta-analysis.

5. Population: all patients with any clinical condition.

6. Intervention: non-activity trackers or drugs 7. Comparison group: all types of comparators 5. Outcome: weight.

8. Human study of any age group and sex.

For detailed information regarding the search strategies please, refer to appendix II.

VI.2.5 Statistical analysis

The methodological quality of the included systematic reviews was assessed using the AMSTAR-2 tool (Shea *et al.*, 2017b). AMSTAR-2 is a recognized and reliable 16-question tool for evaluating the methodological quality of systematic reviews of health care treatments (Shea *et al.*, 2017b; Lorenz *et al.*, 2019).

After assessing the methodological quality of the included systematic reviews using the sixteen AMSTAR-2 questions, the methodological quality rating of each study was determined by calculating the number of "yes" answers to the sixteen AMSTAR-2 questions, as this strategy has been used in numerous related studies (Pieper *et al.*, 2018). To test the hypothesis, we first test the normal distribution of the data between the two groups using the Shapiro Wilk test with a 95% confidence intervals and the P-value was used to decide if the data follows a normal distribution.

After checking the normal distribution of the data between the two groups, a two sample t-test was used to test the hypothesis.

VI.3 Results

VI.3.1 Search and screening results

The search identified 67 documents, which resulted in 20 eligible systematic reviews that provided a meta-analysis of the effect of ICD on mortality in heart failure patients. Subsequently, a search was conducted to identify non-DBMS or drug-based interventions that shared the same populations and outcomes as the ICD studies. Please refer to appendix II for further details.

Following screening and checking based on the inclusion criteria, 14 systematic reviews were also selected as wearable activity trackers out of 44 identified studies. For more information, refer to appendix II. Ultimately, 20 ICD-based and 14 wearable activity tracker-based systematic reviews with their drug-based peers were included in the final data analysis and their methodological quality was assessed.

VI.3.2 Methodological quality

The mean overall grade of methodological quality for digital biomarkers was 9.80 (SE = 0.41, 95% CI = 8.95-10.63), while the mean overall grade of methodological quality for drug-based systematic reviews was 9.76 (SE = 0.33, 95% CI = 9.1-10.43). Refer to table 5 for more information.

Table 5. Mean methodological quality of digital biomarker-based and non-digital biomarker or pharmaceutical intervention systematic reviews

Type of systematic review	Mean	Standard error	95% confidence intervals
Digital biomarkers	9.80	0.41	8.95 -- 10.63
Pharmaceuticals or non-digital biomarkers	9.76	0.33	9.1 – 10.43

VI.3.3 Normal distribution

To assess the normality of the digital biomarkers and non-digital biomarkers or drug-based systematic reviews' methodological quality, the Shapiro-Wilk W test was performed. For the digital biomarkers group, the test revealed a W value of 0.98 ($p = 0.80$), indicating that the distribution of overall grades in this group does not significantly depart from normality. Similarly, for the other group, the W value was 0.95 ($p = 0.13$), confirming that the distribution of overall grades in this group is also not significantly non-normal. For further information, refer to the following table.

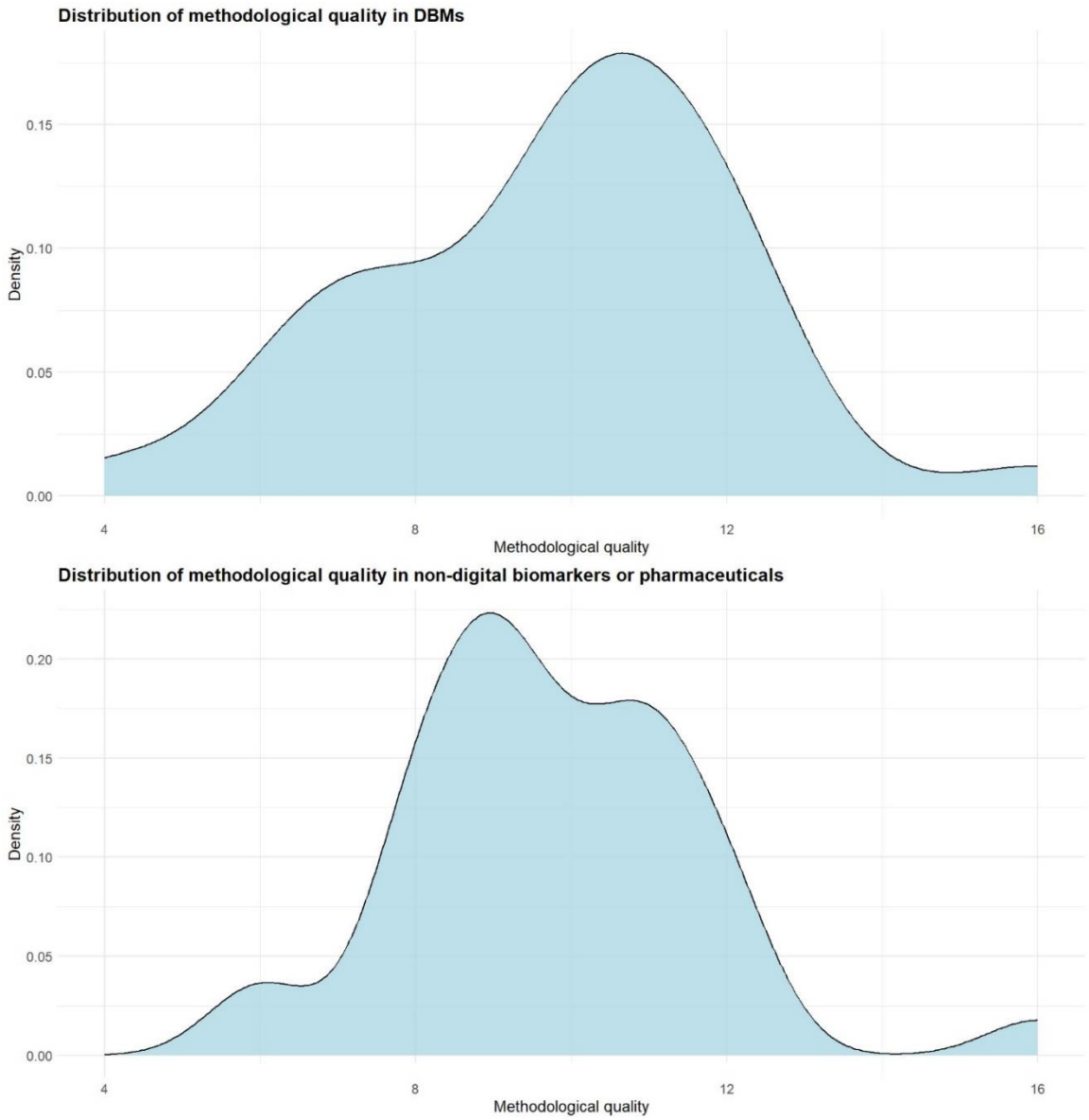


Figure 9. Normal distribution of methodological quality in digital biomarkers and non-digital biomarkers or pharmaceuticals systematic reviews

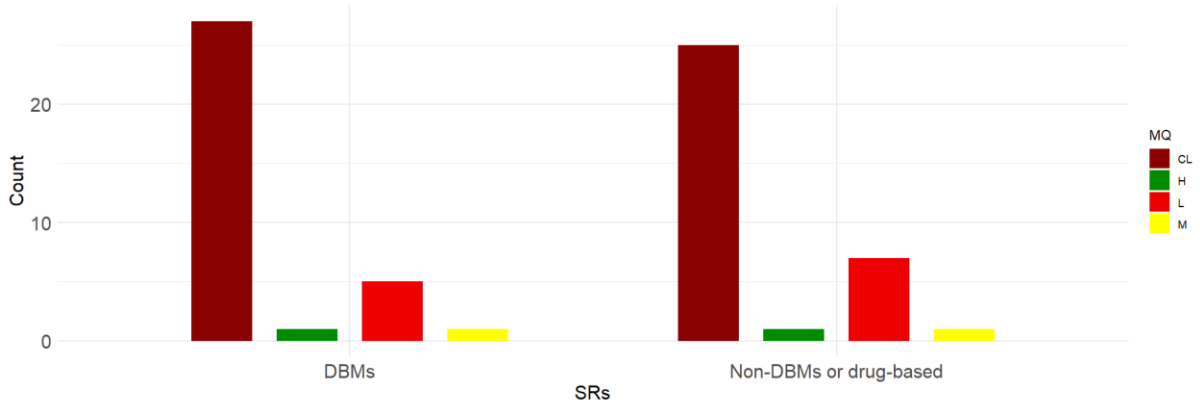
VI.3.4 Statistical analysis

According to the normal distribution testing results, a two-sample t-test was conducted to compare the mean methodological quality of the two groups. The results revealed that the difference between the means was 0.15, showing that the digital biomarkers had a slightly higher mean methodological quality than the non-digital biomarker group. The test statistic, t = 0.29 with 66 degrees of freedom, was calculated, along with the p-value of 0.61. This suggests that there is no statistically significant difference between the mean methodological qualities of the two groups at the 0.05 significance level. Therefore, the null hypothesis that the

mean methodological quality for the digital biomarkers and non-digital biomarkers or drug-based systematic reviews groups are the same cannot be rejected. According to the results, we cannot accept the hypothesis that the methodological quality of digital biomarker-based systematic reviews is significantly lower than that of non-digital biomarker or pharmaceuticals.

VI.3.5 Additional analyses

In order to conduct further analysis, the AMTAR website was employed to evaluate the overall methodological quality of the systematic reviews. Based on the aforementioned analyses, no statistically significant disparity in methodological quality was observed between systematic reviews focused on digital biomarkers compared to those cantered on non-digital biomarkers or drug interventions. Subsequent assessments also revealed that the majority of included systematic reviews in both groups exhibited critically low levels of methodological quality, as depicted in figure 10. For additional details pertaining to the methodological quality of the included systematic reviews, please refer to Appendix III.



SRs	H	M	L	CL
DBMs	1	1	5	27
Non-DBMs or drug-based	1	1	7	25

MQ: Methodological quality, CL: Critically low methodological quality, L: Low methodological quality, M: Moderate methodological quality, H: High methodological quality.

Figure 10. The methodological quality of the included SRs using the AMSTAR-2 tool

VI.4 Discussion

The purpose of this chapter was to investigate whether the methodological quality of digital biomarker-based systematic reviews, identified through PubMed, is significantly inferior to

non-digital biomarker or pharmaceuticals. The study examined thirty-four digital-biomarker-based systematic reviews alongside thirty-four non-digital biomarker or drug-based systematic reviews for comparative analysis. The results indicated that there was no statistically significant difference between the mean methodological qualities of the two groups.

The findings of this study provide confirmatory evidence that the methodological quality of systematic reviews of digital biomarkers does not differ significantly from that of non-digital biomarkers or pharmaceuticals. The field of pharmaceutical research boasts a longstanding historical foundation and a well-regulated framework that governs the conduct of clinical trials and the evaluation of drug efficacy and safety. Stringent regulatory processes, overseen by regulatory bodies such as the Food and Drug Administration (FDA) (FDA, 2022) and the European Medicines Agency (EMA) (EMA, 2020), contribute to the establishment of high-quality standards in systematic reviews pertaining to drug-related research. Conversely, digital health technologies, including the relatively new domain of digital biomarkers, have gained considerable prominence in recent years, promising innovative prospects for healthcare delivery. Given its nascent status, research within this domain is characterized by rapid advancements, diverse study designs, and varying degrees of regulatory oversight. Furthermore, the findings also indicate that researchers and reviewers in both domains operate in a comparable manner, highlighting the absence of significant differences between them.

Besides, the findings revealed that although there was no discernible distinction in the methodological quality of systematic reviews of digital biomarkers and non-digital biomarkers or pharmaceuticals, a prevailing observation across both groups was the prevalence of critically low methodological quality. This suggests that caution should be exercised when relying on these systematic reviews to furnish an accurate and comprehensive evidence synthesis of the existing body of literature (Shea *et al.*, 2017a). Furthermore, the findings from our preceding chapter have revealed a deficiency in statistical power among digital biomarker studies. Consequently, the significance of high-quality systematic literature reviews (SLRs) becomes even more crucial. However, these studies suffer from inherent limitations in methodological quality as was discovered in this chapter. In the presence of smaller-scale studies characterized by high methodological quality, there exists an opportunity for collaboration among smaller companies to generate robust evidence supporting the integration of their technologies into clinical settings. Conversely, when systematic literature reviews (SLRs) exhibit critically low methodological quality, the prospects of such collaborative efforts and subsequent market acceptance are diminished, thereby impeding the timely adoption of these technologies in

medical practice. Several influential factors contribute to the critically low methodological quality observed in systematic reviews encompassing both digital biomarkers and non-digital biomarkers, necessitating consideration from researchers, reviewers, and regulators within both domains. An imperative aspect entails conducting a comprehensive search across multiple electronic databases to ensure the inclusion of all available literature, as only a limited number of systematic reviews in both groups met the criteria for fulfilling this particular requirement outlined in the AMSTAR-2 tool. Inadequate search strategies may introduce bias, as incomplete searches may overlook numerous relevant studies, consequently depreciating the validity of the results and may lead to publication bias (Murad *et al.*, 2018). This importance is accentuated within the realm of digital biomarkers, where the absence of a unified and well-established definition necessitates careful consideration of search strategies (Zrubka *et al.*, 2021).

A notable AMSTAR-2 criterion that a majority of the systematic reviews included in both the digital biomarkers and non-digital biomarkers domains failed to satisfy is the consideration of the impact of risk of bias in the included studies and its influence on the review's findings (Shea *et al.*, 2017a). The inclusion of studies with a high risk of bias can potentially undermine the validity of the results obtained in these reviews (Frampton *et al.*, 2022). Addressing this concern is of paramount importance within the context of systematic reviews, yet a substantial portion of the systematic reviews examined in this study did not meet this criterion. In the existing literature, several studies have assessed the methodological quality of systematic reviews in various areas of digital health. However, to date, no study has directly compared the methodological quality of systematic reviews of digital health interventions with that of pharmaceuticals. For example, several published articles have reported low or critically low methodological quality in various areas of digital health, including the effect of telemonitoring therapies in patients with chronic obstructive pulmonary disease (COPD) (Li *et al.*, 2020), digital health interventions (DHIs) in palliative care (Finucane *et al.*, 2021), e-health and m-health interventions aimed at changing weight, physical activity, alcohol consumption, and smiling behavior (Jamshidi, Heyvaert and Van den Noortgate, 2017; Gold *et al.*, 2021), as well as digital mental health (Villarreal-Zegarra *et al.*, 2022) and implantable cardiac defibrillator-related infection (Motahari-nezhad, Miribonjar and Sadeghdaghighi, 2022). A limited number of studies have reported on the methodological quality of systematic reviews in digital health interventions. These studies have revealed that most of the systematic reviews exhibit moderate or high methodological quality, which is inconsistent with the results of our research. For

example, a systematic review assessing the methodological quality of systematic reviews on e-health technologies for vaccination practices found that a majority of the reviews exhibited moderate methodological quality (Dumit *et al.*, 2018). The assessment of methodological quality of systematic reviews of pharmaceuticals has been carried out in several studies, which have reported similar findings. These studies have shown that a significant proportion of systematic reviews of pharmaceuticals have low and critically low methodological quality, as demonstrated by studies on traumatic dental surgery (Magno *et al.*, 2020), childhood caries (Soares *et al.*, 2021), and chronic pain relief (Bussadori *et al.*, 2020).

VI.4.1 Strengths and limitations

The study has several limitations that need to be acknowledged. Firstly, the study only focused on two types of digital biomarkers (i.e., implantable cardiac defibrillators and wearable activity trackers) and their methodological quality compared to non-digital biomarkers or drug-based interventions. The results may not be generalizable to other types of digital biomarkers or medical interventions. Secondly, the study only included systematic reviews identified through PubMed, which may not be representative of all available systematic reviews on the topics. The present study boasts several strengths. To our knowledge, this is the first study to examine the methodological quality of systematic reviews cantered on digital biomarkers in comparison with their non-digital or pharmaceutical counterparts.

VI.4.2 Conclusion

The development of rigorous standards in pharmaceutical development (FDA, 2023) has likely influenced the creation of guidelines for systematic reviews of digital health. These guidelines commonly incorporate stringent criteria for study design, participant selection, outcome measures, and statistical analysis, ensuring that systematic reviews of pharmaceuticals are executed using meticulous methodology. Conversely, the field of digital biomarkers is still in its early stages and is rapidly evolving, leading to potential variations in the methodological quality of systematic reviews (Iqbal and Biller-Andorno, 2022). Unlike pharmaceuticals, which possess well-established standards for study design and reporting (FDA, 2023), the field of digital biomarkers is still establishing best practices for conducting studies and synthesizing evidence, while the findings of this study demonstrate that there is no significant difference in the methodological quality of systematic reviews between pharmaceuticals and digital biomarkers. Secondly, the distinct characteristics of digital biomarkers may pose unique challenges for conducting systematic reviews. Digital biomarkers can be derived from a multitude of sources, such as wearables (Ferguson *et al.*, 2022), implantables (Anantha

Narayanan *et al.*, 2017), portable devices, mobile applications, and other digital platforms (Motahari-Nezhad, Fgaier, *et al.*, 2022). These sources may differ in their data quality, accuracy, and reliability, which can make it challenging to synthesize the evidence in a meaningful and comprehensive manner (Mukherjee *et al.*, 2022). Furthermore, the nature of digital health interventions and biomarkers makes blinding in clinical studies difficult or impossible to achieve (Monaghan *et al.*, 2021). As a result, there may be inconsistencies in the quality of studies and data sources, which can affect the overall methodological quality of systematic reviews of digital biomarkers. Nonetheless, our study found no significant difference in the methodological quality of systematic reviews between digital biomarkers and pharmaceuticals.

While methodological quality is an essential component of systematic reviews, it is not the only factor that needs to be considered when evaluating the potential clinical utility of digital biomarkers or pharmaceuticals. Assessing the quality of evidence and the cost-effectiveness of these interventions are critical next steps to inform clinical decision-making. In addition to the quality of evidence, cost-effectiveness is also a crucial consideration when evaluating the potential clinical utility of digital biomarkers and pharmaceuticals. While both interventions may show promising results in terms of clinical efficacy, the cost of implementing these interventions may also need to be considered. In some cases, the cost of implementing a digital biomarker or pharmaceutical may be prohibitive, particularly in resource-limited settings. Therefore, it is important to consider the quality of evidence and the cost-effectiveness of digital biomarkers and pharmaceuticals in addition to methodological quality when making clinical decisions. This comprehensive approach to evaluation can help ensure that medical professionals make informed and effective decisions that optimize patient outcomes while minimizing costs.

VII. ASSESSING AND COMPARING THE QUALITY OF EVIDENCE OF DIGITAL BIOMARKER-BASED META-ANALYSES WITH THAT OF NON-DIGITAL BIOMARKERS OR PHARMACEUTICALS

VII.1 Introduction

Evidence-based medicine is an essential part of today's practice. It is so important that it is impossible to imagine contemporary health care if evidence and its quality are neglected (Szajewska, 2018b). SRs and meta-analyses give a less biased, more exact estimate on a clinical problem, making them the gold standard in evidence-based medicine (Oxman, 1993).

In addition to the many advantages of meta-analyses for medical research, some disadvantages should also be noted. For example, when the studies included in a meta-analysis are so heterogeneous, not enough studies and equal sizes in a meta-analysis, the exclusion of some other related studies from a meta-analysis, the inclusion of studies with different research questions, all these are reasons that affect the validity of a meta-analysis (Lee, 2018). In this regard, there is a validated tool, namely GRADE (Grading of Recommendations Assessment, Development, and Evaluation Working Group), to assess the validity of the results of a meta-analysis (quality of evidence). As discussed in previous chapters, the number of digital biomarker-based systematic reviews is increasing. Therefore, we aim to compare the quality of evidence from meta-analyses based on digital biomarkers with the quality of evidence from non-digital biomarkers (e.g., drug therapies) to determine which of them provide high quality evidence. Assessing the quality of evidence from digital biomarker-based meta-analyses compared with non-digital biomarkers (e.g., pharmaceuticals) will help physicians and health policy makers select the best treatment strategies.

VII.2 Methods

VII.2.1 Search strategies and inclusion criteria

We utilized the identical set of studies employed in the previous hypothesis to investigate the current research question.

VII.2.2 The assessment of the quality of evidence

We assessed the quality of evidence for each outcome using the GRADE system (Guyatt *et al.*, 2008; Schünemann *et al.*, 2013). By default, GRADE classifies evidence from randomized controlled trials as high quality. However, this rating can be downgraded based on the

assessment of the following five quality domains: 1) risk of bias (Guyatt, Oxman, Vist, *et al.*, 2011), 2) inconsistency (Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Glasziou, *et al.*, 2011), 3) imprecision (Guyatt, Oxman, Kunz, Brozek, *et al.*, 2011), 4) publication bias (Guyatt, Oxman, Montori, *et al.*, 2011), and 5) indirectness (Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Falck-Ytter, *et al.*, 2011). Depending on the severity of the quality concerns, a downgrade of 0, 1, or 2 can be proposed for each domain.

VII.2.2.1 The risk of bias

Risk of bias is the problems with the design and execution of individual studies of healthcare interventions (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a). The risk of bias was be evaluated based on the following criteria: if 75% or more of the included studies had a low risk of bias for a particular outcome, no downgrading was imposed. If less than seventy-five percent of the included studies had a low risk of bias, or if the risk of bias was not stated, one downgrade was applied. (Pollock *et al.*, 2016).

VII.2.2.2 Inconsistency

Inevitably, differences exist across the research included in a systematic review. In a systematic review, heterogeneity refers to any kind of variation between researches. It might be useful to differentiate between several forms of heterogeneity (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a). The stated heterogeneity for each outcome was used to evaluate inconsistency. If the I^2 statistic was less than or equal to 75%, there would be no downgrade. If the I^2 statistic exceeded 75%, a downgrading was awarded. If just one research was considered for the outcome, there was no downgrading. A downgrading was issued if heterogeneity was not noted. (Pollock *et al.*, 2016).

VII.2.2.3 Imprecision

In general, results are imprecise when studies include a small number of patients and events, resulting in a wide confidence interval (CI) around the estimate of effects. As a result of the ambiguity of the results, the quality of the evidence may be judged to be worse than it otherwise would have been (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a). We assessed Imprecision by evaluating the sample size (Zhang *et al.*, 2019). If the total sample size exceeds 2,000, the evidence was lowered.(Schünemann *et al.*, 2013). If the total sample size was less than 200, we degraded the rating by one point. Using STATA 16 and power analysis, we determined the optimum information size (OIS) for a pooled sample size between 200 and 2000 as follows: (Schünemann *et al.*, 2013): assuming a weak effect size (Cohen, 1988), We computed the sample size for a randomized controlled trial assuming a balanced

sample, 0.8 power, and 0.05 significance level. When the computed sample size was bigger than the pooled sample size, one downgrade was applied. (Schünemann *et al.*, 2013; Zhang *et al.*, 2019). For the small effect size, the following technique was utilized: a Cohen's d of 0.2 for continuous measurements and 1.68 for the odds ratio. A small effect size of 1.68 was also predicted for the risk ratio and hazard ratio, assuming a prevalence of 0% in the unexposed population. (Cohen, 1988; Higgins *et al.*, 2019a).

VII.2.2.4 Publication bias

Publication bias is the failure to publish research results based on the direction or significance of the study's findings (DeVito and Goldacre, 2019). Using the trim-and-fill approach provided by Duval and Tweedie, the possible influence of publication bias on effect size estimations was evaluated for each outcome. (Duval and Tweedie, 2000). Potentially missing studies were imputed, and the impact size of the whole data set was reassessed. If the imputation affected the findings of the analysis (for example, a significant effect size became non-significant or the magnitude of the effect size changed), we would degrade the study owing to publication bias. (Duval and Tweedie, 2000). Due to the low power of risk of bias tests when applied to fewer studies, we examined publication bias only in meta-analyses comprising at least 10 studies, per Cochrane Handbook guidelines. (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a).

VII.2.2.5 Indirectness

In a systematic review, when a meta-analysis compares the effect of two different types of interventions on the same outcome or evaluates the effect of one intervention on two different outcomes, there is immediacy in a meta-analysis (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022a). When evaluating indirectness for each outcome, we took into account differences between the included studies and the meta-research analysis's question. (Terrin, Schmid and Lau, 2005). If the population, treatments, or comparators of the studies did not align with the primary goals of the meta-analysis, a downgrading of 1 or 2 was considered, depending on the degree of this misalignment.

VII.2.2.6 The overall quality of evidence

The overall assessment of the quality of evidence for each outcome was based on the recommendations of Pollock et al. (Pollock *et al.*, 2016). Evidence was rated as high quality if it was extremely unlikely that further research would affect our confidence in the effect estimate (0 downgrades); as moderate quality if further research was likely to have a significant effect on our confidence in the effect estimate and could change the estimate (1-2 downgrades);

as low value if further research would likely have a significant impact on our confidence in the effect estimate and could change the estimate (three to four downgrades); and as very low value if an effect estimate was very unclear. (5-6 downgrades) (Guyatt *et al.*, 2008; Pollock *et al.*, 2016).

VII.2.3 Statistical analysis

Meta-analyses were divided into two groups: digital biomarkers and non-digital biomarkers or pharmaceuticals, whose quality of evidence was assessed using the above strategies. If the quality of evidence of a meta-analysis was low or very low, we selected "yes". Otherwise (if the quality of evidence of a meta-analysis was high or moderate), the quality of evidence was considered as "no". Consequently, the Fisher exact test using 95% confidence intervals was applied to test the hypothesis.

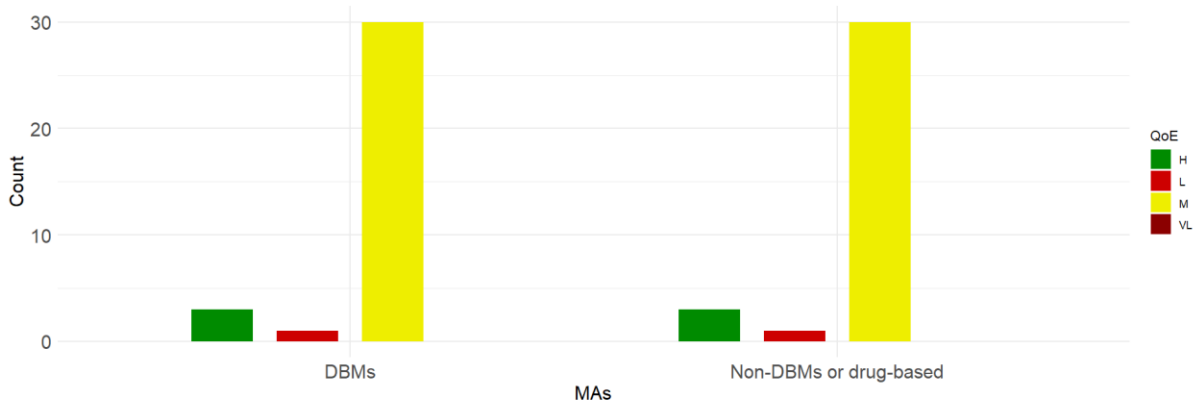
VII.3 Results

VII.3.1 Search and screening results

A total of 34 meta-analyses focusing on DBMs and 34 meta-analyses pertaining to non-DBMs or pharmaceuticals were included in the final analysis. For further details about the characteristics of the included and excluded systematic reviews, refer to appendix II.

VII.3.2 Quality of evidence

Appendix IV shows the quality of evidence assessment results. Figure 11 also shows the level of evidence of each type of intervention. The present analysis incorporated 34 digital biomarker-based meta-analyses and 34 non-digital or drug-based meta-analyses, as indicated by the table. The results of the investigation revealed an equal distribution of evidence between the two interventions, with each modality exhibiting three instances of high-quality evidence, thirty instances of moderate-quality evidence, and a low-quality evidence. Moreover, no evidence of very low quality was observed for either digital biomarkers or pharmaceuticals. The majority of the meta-analyses included in both groups were of moderate quality of evidence, with a substantial proportion of 30 out of 34. Conversely, the number of high-quality evidence meta-analyses was limited to 3 out of 34, whereas only 1 out of 34 was rated as low-quality in both groups.



MAs	H	M	L	VL
DBMs	3	30	1	0
Non-DBMs or drug-based	3	30	1	0

QoE: Quality of evidence, H: High quality of evidence, M: Moderate quality of evidence, L: Low quality of evidence, VL: Very low quality of evidence.

Figure 11. Quality of evidence of MAs according to their intervention type

VII.3.3 Statistical analysis

In order to evaluate the hypothesis that the ratio of digital biomarker-based meta-analyses exhibiting low or critically low quality of evidence surpasses that of drug or non-digital biomarker-based meta-analyses, the Fisher exact test was implemented. The results show that the Fisher's exact test has a value of 1.000, which indicates no evidence of an association between the quality of evidence of digital biomarkers and drug-based meta-analyses. The Fisher's exact test, which tests the hypothesis that one group has higher proportion than the other group, has a p-value of 0.97, which also indicates no evidence of an association. Therefore, based on the observed data, we can conclude that there is no significant difference between these two groups and we can reject the hypothesis that the proportion of meta-analyses of digital biomarker-based interventions with low or very low quality evidence is higher than that of non-digital biomarkers or drug-based interventions.

VII.3.4 Additional analyses

VII.3.4.1 Heterogeneity, sample size, number of included studies, and risk of bias

Supplementary analyses were conducted to compare heterogeneity, sample size, the number of included studies, and the percentage of studies with a low risk of bias between the two groups.

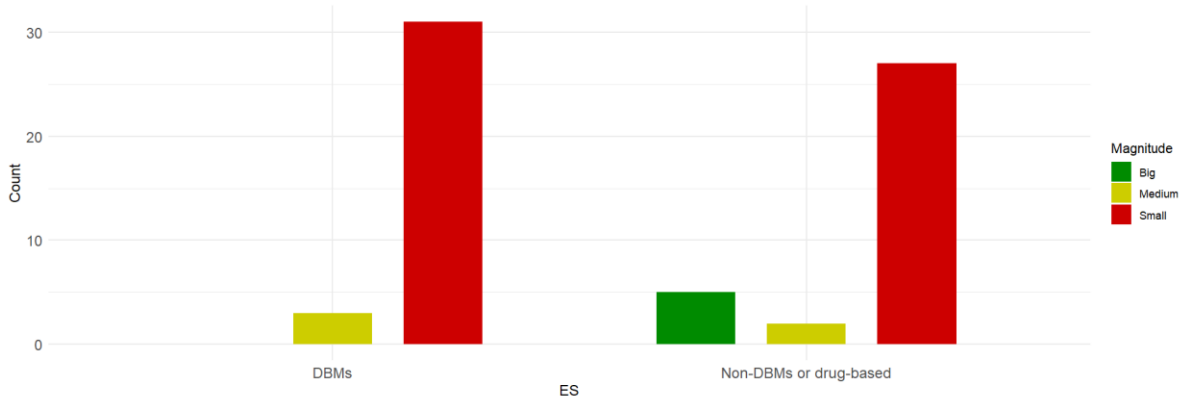
The Shapiro-Wilk test results indicated a non-normal distribution for all the aforementioned variables in both groups (p-values < 0.05). Consequently, a rank sum test was employed to assess the differences, and the findings ultimately demonstrated no statistically significant distinction between the two groups regarding heterogeneity, the number of included studies, sample size, and risk of bias (p-values > 0.05). Table 6 shows the descriptive statistics of the variables in both groups' digital biomarkers and non-digital biomarkers or drug-based meta-analyses.

Table 6. The descriptive statistics of variables affecting quality of evidence in the two groups

Variable	Type of meta-analysis	Mean	Standard error	95% confidence intervals
Heterogeneity	DBMs	29.3	5.72	17.68 – 40.94
	Non-DBMs or pharmaceuticals	26.79	6	14.58 - 39
Sample size	DBMs	3497	925.92	1613.11 – 5380.71
	Non-DBMs or pharmaceuticals	3229.68	614.77	1978.91 – 4480.44
Number of included studies	DBMs	8.29	0.94	6.38 – 10.2
	Non-DBMs or pharmaceuticals	7.29	0.92	5.43 – 9.16
The percentage of included studies with low risk of bias	DBMs	22.35	5.64	10.87 – 33.83
	Non-DBMs or pharmaceuticals	30.43	6.06	18.11 – 42.76

VII.3.4.2 Effect size magnitude

The distribution of effect sizes' magnitudes within both groups is illustrated in Figure 12. A substantial proportion of effect sizes within both groups exhibited a small magnitude of effect. Conversely, while no effect size of digital biomarkers exhibited a big effect, a limited number of non-digital biomarkers or pharmaceuticals yielded big magnitude of effects. For further information, please see Figure 12.



ES	Big	Medium	Small
DBMs	0	3	31
Non-DBMs or drug-based	5	2	27

Figure 12. The distribution of effect size magnitude in DBMs and non-DBMs or drug-based MAs

VII.4 Discussion

Based on the observed data, it was concluded that there is no significant difference between these two groups. Upon analysis, the present study has identified that the majority of outcomes and meta-analyses in both groups including digital biomarkers and pharmaceuticals exhibit a moderate quality of evidence. This observation is in line with numerous similar studies published in the domain of digital health technologies, which have consistently reported comparable findings. For instance, Du et al. 2020 reported moderate quality of evidence for e-health interventions in chronic low back pain, while previous studies investigating the efficacy of pacemakers (Psaltikidis, Costa and Graziano, 2021) and the use of sensors to prevent falls in hospitalized patients (Cortés *et al.*, 2021) have also documented moderate quality of evidence. In a recently published article, the quality of evidence of digital biomarkers was evaluated and the researchers found that most digital biomarkers have moderate quality of evidence (Motahari-Nezhad, Al-Abdulkarim, *et al.*, 2022b). On the contrary, there exist non-digital biomarker studies that yield somewhat comparable results, such as investigations into the impact of envelope on cardiac infections (Motahari-nezhad, Miribonjar and Sadeghdaghighi, 2022). Furthermore, several studies have reported varying degrees of evidence quality in both digital and non-digital health domains. For instance, Lynch et al. 2019 documented that wearable activity monitors and smartphone apps generate evidence of low

and very low quality in measuring step count and physical activity in the digital biomarker domain. Similarly, Stevenson et al. 2019 rated the evidence quality of e-health interventions in death, weight, and sodium intake in chronic kidney disease as low. Based on the literature, e-health interventions produce low-quality evidence in the disability of chronic low back pain (Du *et al.*, 2020). Additionally, studies examining the effectiveness of e-health treatments in weight loss (Lau *et al.*, 2020), digital shared decision-making (Vitger *et al.*, 2021) and e-health interventions for rheumatoid arthritis patients (L. Zhou *et al.*, 2022) all found the evidence to be of low or very low quality.

There was no statistically significant differences between the two groups concerning heterogeneity, sample size, the number of included studies, and the risk of bias associated with the included studies. These findings suggest that key factors such as study design, participant characteristics, and intervention characteristics do not significantly differ between the two intervention types (Sedgwick, 2015). It is noteworthy that despite non-digital biomarkers or pharmaceuticals having a more robust regulatory framework in clinical research (Hoebert, 2015), no substantial difference in heterogeneity was observed, indicating a research gap in this specific domain. Additionally, it is important to highlight that the mean heterogeneity for both groups was below the critical value of 50%. This is of particular importance, as heterogeneity exceeding 50% can potentially undermine the validity of the results (Sedgwick, 2015). Another noteworthy aspect pertains to the risk of bias present in the studies included in the meta-analyses of both groups, which demonstrated a non-significant difference. Although the observed difference did not reach statistical significance, the findings indicate that, on average, a limited number of studies with low risk of bias were included in the meta-analyses for both groups. These results suggest that the majority of included studies may possess certain degrees of bias. While the lack of statistical significance implies that the difference in bias between the two groups may not be substantial, it is crucial to acknowledge the presence of studies with a potentially suboptimal methodological rigor across both groups. These findings emphasize three key considerations that merit careful attention. Firstly, the results imply a pressing need for more robust clinical studies conducted with rigorous methodologies, devoid of any potential biases. Such studies are essential to facilitate the aggregation of reliable and validated results in the form of meta-analyses. Addressing this need would contribute to enhancing the quality and validity of synthesized evidence. Secondly, there is a clear requirement for the establishment of more concrete regulations governing clinical studies. Strengthening the regulatory framework would play a pivotal role in improving the overall

status of bias in clinical research. By implementing comprehensive guidelines and standards, the aim is to minimize potential sources of bias, thereby elevating the credibility and integrity of study outcomes. Furthermore, it is noteworthy that during the assessments conducted, a small number of meta-analyses within both the digital biomarkers and non-digital biomarkers (with only one instance in each group), as well as pharmaceuticals, failed to evaluate and adequately report the risk of bias associated with the included studies. This observation highlights a research gap and emphasizes the necessity for reviewers and clinical research regulators to be vigilant in recognizing and addressing this shortcoming. These considerations emphasize the significance of conducting high-quality, unbiased clinical studies, implementing robust regulatory measures, and promoting rigorous reporting standards within meta-analyses. By addressing these aspects, we can enhance the quality and credibility of synthesized evidence in the field of clinical research.

Digital biomarkers and pharmaceuticals had similar levels of evidence supporting their use in clinical practice, as the results indicated. The results showed that both groups digital biomarkers and pharmaceuticals have moderate quality of evidence, meaning that the true effect is probably close to the estimated effect (Schünemann *et al.*, 2013). As outlined in the preceding chapter, there is no discernible difference in methodological quality between the two treatment groups, digital biomarkers and pharmaceuticals, which confirms that they generate equivalent levels of evidence. Moreover, the data analysis confirmed that both intervention groups produce generally small effect. However, it is important to note that while pharmaceuticals adhere to a well-established framework and regulations to be incorporated into the healthcare system, digital biomarkers are a relatively new area that is expanding and penetrating rapidly, but their managerial frameworks and regulations have not yet been fully established. Consequently, due to the inherent differences between these two treatment options, managers should consider implementing well-established rules when selecting between them. When clinicians are deciding between digital biomarkers and pharmaceuticals, since both create the same level of evidence and effect, neither should be prioritized over the other. Instead, the cost-effectiveness of these treatments should be considered, and future studies should focus on exploring this aspect.

VII.4.1 Strengths and limitations

The use of the GRADE tool to assess the quality of evidence added a transparent approach to the evaluation process. However, the study was limited to a single database (PubMed) and may not represent the entire landscape of digital biomarker-based and drug-based interventions. The

analysis was limited to two types of digital biomarkers, and the results may not be generalizable to other types of digital biomarkers. Finally, the study did not explore the potential benefits or harms of the interventions, but solely focused on the quality of evidence, which may not provide a comprehensive picture of the effectiveness and safety of the interventions.

VII.4.2 Conclusion

Both digital health technologies and pharmaceutical interventions contribute equally to the body of evidence. Furthermore, it is crucial to address the issue of risk of bias as a significant factor affecting the reliability of research outcomes. To overcome this limitation, it is essential to conduct more clinical studies with a low risk of bias. By ensuring that bias is minimized or eliminated, researchers can generate unbiased outcomes that can be utilized in meta-analyses.

VIII. PUBLICATION BIAS IN DIGITAL BIOMARKER-BASED META-ANALYSES AND NON-DIGITAL BIOMARKER-BASED OR PHARMACEUTICAL META-ANALYSES.

VIII.1 Introduction

Publication bias may have a detrimental impact on the validity of effect size findings from meta-analyses. Publication bias may lead to an incorrect pooled estimate of a treatment effect in a meta-analysis (Almalik, Zhan and van den Heuvel, 2021). For this reason, the results of a meta-analysis are only as reliable as the data that support them; for example, including only published studies could lead to an exaggeration of the effectiveness of digital biomarker interventions, whereas including unpublished studies with insignificant results could lead to a shift in the mean effect estimate. Therefore, we performed a systematic review of systematic reviews to evaluate publication bias in digital biomarker meta-analyses. Due to the fact that digital health technologies and digital biomarkers lack unified definitions and names compared to pharmaceuticals, a comparative study was also conducted to compare the publication bias in digital biomarkers and non-digital biomarkers. It was hypothesized that publication bias is more prevalent in digital biomarker-based meta-analyses than in non-digital biomarker-based meta-analyses. This study assessed the publication bias difference between meta-analyses of digital biomarker-based interventions and non-digital biomarker-based interventions.

VIII.2 Methods

This section is based in part on an our article published in the Journal of Medical Internet Research (Hossein Motahari-Nezhad, Al-Abdulkarim, et al. 2022). According to the results of this article, 22 systematic reviews with 95 meta-analyses (outcomes) were included. According to the Cochrane Handbook, the assessment of publication bias should be performed in meta-analyses with at least ten studies. Therefore, meta-analyses with at least ten studies were considered for the assessment of publication bias.

VIII.2.1 Inclusion criteria, search strategy, and screening of digital biomarker-based meta-analyses

The methods explained in hypothesis II were used to identify digital biomarker-based meta-analyses.

VIII.2.2 Inclusion criteria, search strategy, and screening to identify non-digital biomarker-based meta-analyses

To compare the publication bias of digital biomarker-based meta-analyses with that of non-digital biomarkers or pharmaceuticals, each identified digital biomarker meta-analysis was considered and the population and outcome were taken into account. Consequently, in the same year, a similar search with population and outcome was conducted to find a meta-analysis with similar population and outcome but with a non-digital biomarker intervention. If we did not find a similar meta-analysis with non-digital biomarkers in the same year, we expanded the search to other years. Within the realm of meta-analyses based on wearable activity trackers, the reported outcomes primarily encompassed physical activity, moderate to vigorous physical activity or step count. In order to identify corresponding meta-analyses pertaining to non-DBM-based interventions or pharmaceutical interventions, an exploration was conducted to identify analogous outcomes, such as weight.

VIII.2.3 The assessment of Publication bias

Twenty meta-analyses including at least 10 studies of interventions based on digital biomarkers were identified, and similar searches were conducted to find other similar meta-analyses with the same population and outcomes but a non-digital biomarker or pharmaceutical intervention. After another 20 meta-analyses on non-digital biomarkers were found, the trim-and-fill method was used to:

1- Identify the number of missing studies in meta-analyses
2- Determine the change in effect size of each meta-analysis, called adjusted effect size, using the trim-and-fill method, and finally the meta-analyses of the two groups were compared in terms of the number of missing studies and adjusted effect sizes.

The 'trim and fill' technique seeks to detect and adjust for publication-biased funnel plot asymmetry (Duval and Tweedie, 2000). The basic idea is to (1) "trim" (eliminate) the smaller studies responsible for the asymmetry in the funnel plot, (2) use the trimmed plot to estimate the real "center" of the funnel, and (3) reinstate the missing studies and their "counterparts" around the estimated center (filling). An adjusted intervention effect is produced through a meta-analysis that includes the missing studies, and an estimate of the number of missing studies is also provided (Duval and Tweedie, 2000). The trim and fill technique estimates the number of missing studies and the intervention impact "adjusted" for publication bias (Duval and Tweedie, 2000).

VIII.2.4 Statistical analysis

In this study, we tested two following hypotheses:

1. As the first hypothesis, it was hypothesised that the proportion of missing studies in meta-analyses based on digital biomarkers would be significantly higher compared with that of non-digital biomarkers or pharmaceuticals.

2. As a second hypothesis, it was hypothesised that the difference between the reported effect size and the adjusted effect size in meta-analyses of digital biomarkers would be significantly higher than that of non-digital biomarkers or pharmaceuticals.

After finding the final meta-analyses that meet the eligibility requirements, we conducted meta-analyses for each outcome using Stata 16 software. To rerun the meta-analysis, the effect sizes and confidence intervals of each research included in the meta-analyses were utilized. The trim-and-fill approach determined the number of missing studies and the adjusted effect size. In order to establish the level of publication bias, the reported effect sizes and the effect sizes recalculated using the trim-and-fill approach (adjusted effect size) were compared. The same effect model was utilized as in the original meta-analyses (fixed effect and random effect). We conducted a random-effects meta-analysis using the DerSimonian and Laird model.

The assessment of the normal distribution for both the number of missing studies and the discrepancy between the reported and adjusted effect sizes was conducted using the Shapiro-Wilk test. Subsequently, a non-parametric Mann-Whitney U test was utilized to investigate potential significant differences between the two groups concerning the number of missing studies and the magnitude of the effect size change. Regarding the second hypothesis of this chapter (the magnitude of the change in effect sizes between the two groups), the trim-and-fill method was first used to calculate the adjusted effect size by importing missing studies into the meta-analyses. Then, the relative change between the reported and adjusted effect sizes was measured using the following formula:

$$RC_{ES} = \frac{ES_{adjusted} - ES_{reported}}{\left|ES_{reported}\right|}$$

$ES_{reported}$ presents the effect size calculated in the meta-analysis. $ES_{adjusted}$ denotes the adjusted effect size calculated using the trim and fill method.

VIII.3 Results

Hypotheses 5 and 6 aimed to evaluate and compare the meta-analyses of digital biomarkers with those of non-digital biomarkers or pharmaceuticals in relation to publication bias.

VIII.3.1 Search and screening results

Based on the predetermined inclusion and exclusion criteria, a total of 25 systematic reviews consisting of 95 meta-analyses focusing on digital biomarkers were identified and finally 20 meta-analyses from 13 SRs which includes at least 10 studies were deemed eligible for the final examination. Figure 6 shows the selection process of studies. Furthermore, supplementary searches were conducted to identify all peer studies related to non-DBMs or drug-based meta-analyses from the aforementioned twenty meta-analyses. Detailed information regarding the characteristics of the included studies and the search strategies employed to identify non-DBMs or drug-based meta-analyses can be found in appendix V.

VIII.3.2 Publication bias

The number of missing studies identified and the change in effect size magnitude for each of the included meta-analyses in percentage are presented in appendix V.

Based on the results of data analysis, the mean number of missing studies in meta-analyses of digital biomarkers and non-DBMs or pharmaceuticals was found to be 2.3 and 2.35, respectively. Furthermore, the range of missing studies varied between 0 to 7 for DBMs and 0 to 9 for pharmaceuticals as can be seen in table 7.

Table 7. Descriptive statistics of the number of missing studies imputed in the meta-analyses

Type of the intervention	Observation	Mean	Standard deviation	Min	Max
DBMs	20	2.3	2.1	0	7
Non-DBMs or Pharmaceuticals	20	2.35	2.43	0	9

Table 8 presents the alterations in effect sizes in both groups as a consequence of imputed studies. Specifically, the relative change in effect size of digital biomarkers exhibited an

average of 0.14, whereas for pharmaceuticals, this alteration was calculated as 0.08. Despite the minimum of relative change being identical for both groups (zero), the maximum of relatice change in effect sizes was higher in digital biomarkers, being approximately twice that of pharmaceuticals (0.72 compared to 0.32).

Table 8. Descriptive statistics of the relative change in the effect sizes in percentage by considering the imputed studies in meta-analyses

Type of the intervention	Observation	Mean	Standard deviation	Min	Max
Digital biomarkers	20	0.14	0.21	0	0.72
Pharmaceuticals	20	0.05	0.08	0	0.32

VIII.3.3 Normal distribution

To determine any significant distinctions between the number of missing studies and alterations in recalculated effect sizes using imputed studies, an initial step was to assess the normal distribution of the dataset. The Shapiro-Wilk test indicated that the number of missing studies in digital biomarkers was normally distributed (p-value>0.05), whereas the missing studies in pharmaceuticals did not follow a normal distribution (p-value<0.05). Additionally, normality testing of changes in effect sizes of meta-analyses confirmed that both groups exhibited a non-normal distribution (p-value<0.05). Figure 13 shows the distribution of the variables in the two groups.

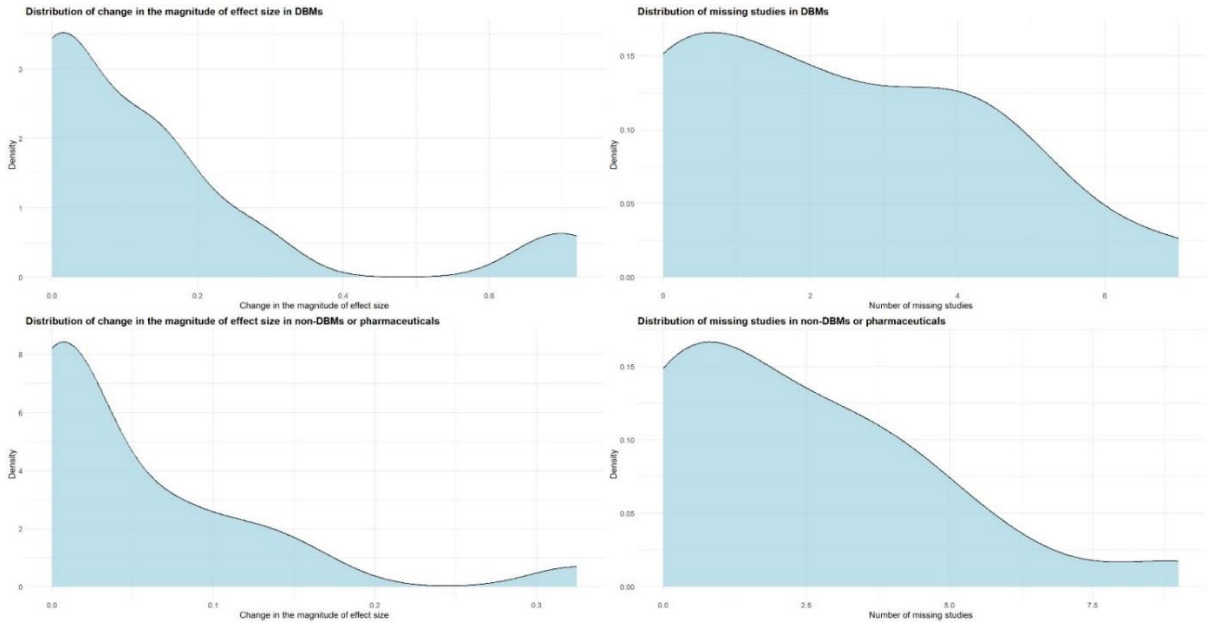


Figure 13. The distribution of the number of missing studies and the change in the effect sizes

VIII.3.4 Data analysis

To examine any significant differences between the two groups concerning the number of missing studies and the alterations in recalculated effect sizes incorporating imputed studies, a non-parametric Mann-Whitney U test was employed. The outcomes of the test failed to reveal any significant disparities between the two groups in terms of either the number of missing studies or the changes in effect sizes (p-value >0.05). Thus, based on the results, we reject the two hypotheses proposing that the proportion of missing studies and the changes in effect size of digital biomarker-based meta-analyses is considerably higher than that of non-DBMs or pharmaceuticals. Therefore, we conclude that no significant distinctions exist between the two groups concerning the number of missing studies and alterations in effect sizes. Further information is provided in Table 9.

Table 9. The results of Mann-Whitney U for testing a difference in missing studies and the change in effect sizes between the two groups

Variable	Type of intervention	Observations	Rank sum	Adjusted variance	P-value
Missing studies	Digital biomarkers	20	414	1312.05	0.91
	Drugs	20	406		
The change in the effect sizes	Digital biomarkers	20	450	1330	0.27
	Drugs	20	370		

VIII.3.5 Additional analysis

Considering the lack of significant difference in the change of adjusted effect sizes between two groups, namely DBMs and non-DBMs or pharmaceutical interventions, an examination was conducted to assess whether publication bias influenced the outcome of the meta-analyses. This assessment involved comparing the reported effect size in the meta-analysis to the adjusted effect size obtained through the trim and fill method. Any alterations in effect size magnitude or the significance of the meta-analysis result were considered indicative of publication bias significantly affecting the meta-analysis result. The findings revealed that publication bias affected the results of meta-analyses involving DBMs in only two instances. Firstly, in a study evaluating the impact of wearable activity trackers on moderate to vigorous physical activity, the initially reported effect size was significant; however, after imputing seven additional studies, the effect size became non-significant (Kirk *et al.*, 2019). Secondly, in another study investigating the effect of Fitbit on weight (Ringeval *et al.*, 2020), the previously significant effect size transformed into a non-significant one**.** Nonetheless, the outcomes of meta-analyses concerning non-DBMs or pharmaceutical interventions remained unaffected by the presence of publication bias.

VIII.4 Discussions

The findings indicate that the average count of missing studies was equivalent in both the DBMs and non-DBMs or pharmaceuticals meta-analyses groups. Moreover, a mere six out of a total of 20 meta-analyses had no missing studies in both digital biomarkers (Hodkinson *et al.*, 2019; Hsin-Yen Yen Yen and Chiu, 2019; Franssen *et al.*, 2020; Ringeval *et al.*, 2020; Wang *et al.*, 2020) and pharmaceuticals (Min *et al.*, 2017; De Menezes *et al.*, 2019; Abdollahi *et al.*,

2020; Hansen *et al.*, 2020; Talenezhad *et al.*, 2020; Janani *et al.*, 2021). Based on the results of the trim and fill method, it can be inferred that the reported effect sizes in these analysed studies are not affected by publication bias, and that overestimation is not present in the corresponding meta-analyses (Yang *et al.*, 2023). The largest count of missing studies in a meta-analysis, amounting to nine studies, was estimated and observed in a pharmaceutical meta-analysis examining the metabolic impact of the dual SGLT 1/2 inhibitor sotagliflozin on body weight (Wu *et al.*, 2022). The meta-analytic analysis of digital biomarkers revealed that the most substantial count of missing studies pertained to a meta-analysis investigating the impact of wearable activity trackers on moderate to vigorous physical activity in individuals with chronic cardiometabolic disorders (Kirk *et al.*, 2019).

The results show that on average, the relative change in effect size was 14% and 5% for DBMs and non-DBMs or pharmaceuticals meta-analyses, respectively. The change in effect size of digital biomarkers was approximately three times that of pharmaceuticals. The most considerable change in digital biomarkers was observed in a meta-analysis involving seven missing studies, where the recalculated effect size altered by 72%. This analysis evaluated the influence of wearable activity trackers on physical activity in patients with cardiometabolic diseases (Kirk *et al.*, 2019). Among the meta-analyses focused on non-DBMs or pharmaceuticals, the greatest change in the recalculated effect sizes was observed in a study investigating the impact of L-carnitine supplementation on weight loss and body composition. Specifically, after the inclusion of six missing studies, the recalculated effect size changed by 32% (Talenezhad *et al.*, 2020). However, in the meta-analysis containing the highest number of missing studies (nine missing studies), the change in effect size was only 12% (Wu *et al.*, 2022).

Based on the findings of this chapter, there was no discernible discrepancy in terms of the number of missing studies and the inconsistency between adjusted and reported effect sizes between digital biomarkers and pharmaceuticals-based meta-analyses and publication bias resulted in change only in the results of two DBMs meta-analyses. The results suggest that while there may be a perception that digital biomarker-based meta-analyses are more susceptible to missing data due to the rapidly evolving nature of digital technology (Cuff, 2023) and the associated challenges in study design (Duffy, Christie and Moreno, 2022), and regulatory gaps (Iqbal and Biller-Andorno, 2022), this does not appear to be the case when compared to pharmaceutical-based meta-analyses. The presence of publication bias in meta-analyses can be attributed to a myriad of factors. However, the quality of the search conducted

during systematic reviews, as delineated in the Cochrane handbook (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022b), appears to be the foremost and crucial determinant among these various factors (Motahari-Nezhad, 2023). To mitigate the likelihood of publication bias, it is recommended that multiple databases and sources be searched, a diversity of search terms and strategies be employed, and unpublished studies and grey literature be included in the analysis (Paez, 2017). This is especially important in digital health, where there is still a research gap in the definition of digital health and challenges in grey literature search that can result in publication bias (Zrubka *et al.*, 2021). When appraising the methodological quality of a systematic review, one of the foremost considerations is the rigor of the search process, as accentuated by the AMSTAR-2 tool (Shea *et al.*, 2017a). In the third hypothesis of this thesis, we evaluated the methodological quality of systematic reviews encompassing digital biomarkers and pharmaceuticals. The findings indicated that a limited proportion of both digital biomarkers and pharmaceuticals fulfilled this criterion of the AMSTAR-2 tool, denoting a significant gap not only in digital biomarkers as a nascent and developing field of research (Cuff, 2023) but also in domain of pharmaceuticals. This result highlights the need for greater attention and emphasis on the search process within systematic reviews, particularly given its vital role in ensuring the completeness and reliability of meta-analyses. Henceforth, both digital biomarkers and pharmaceutical systematic reviews, in their pursuit to manage their study and establish a comprehensive search, must take into account various factors. These factors include conducting searches across multiple electronic databases, exploring the grey literature and different trial registries, performing advanced search queries employing diverse operators, and ultimately, scrutinizing the reference lists of the included studies to elevate the quality of their systematic reviews and mitigate the possibility of publication bias (Motahari-Nezhad, 2023).

VIII.4.1 Strengths and limitations

Before interpreting the results of this chapter, it is important to acknowledge its limitations. Firstly, the search conducted to identify digital biomarker-based meta-analyses was restricted to the years 2019 and 2020, and a separate search was carried out to identify meta-analyses related to pharmaceutical interventions. However, due to differences in outcome reporting, it was not always possible to identify a pharmaceutical intervention with an exact match to the digital biomarker-based outcome. In such cases, an outcome that was similar in nature, such as weight for an outcome of physical activity or number of steps, was selected for pharmaceutical

based meta-analyses. Furthermore, the limited number of included meta-analyses (20 digital biomarker-based and 20 pharmaceutical-based) can also be considered a limitation.

VIII.4.2 Conclusion

Despite the fact that non-digital health or pharmaceutical interventions adhering to more well-established regulations and clinical research practices, including precise definitions and standardized nomenclature, no significant disparities were observed when compared to digital biomarkers. The absence of noteworthy distinctions between the two groups regarding the number of missing studies and relative changes in effect sizes suggests a comparable level of search quality. To ensure comprehensive inclusion of all relevant studies in a meta-analysis, it is essential to conduct a more comprehensive search in both groups, thereby minimizing the occurrence of missing studies across all meta-analyses. Achieving this objective necessitates the establishment of robust guidelines and the formulation of specific definitions and terminologies for digital health technologies within the clinical study domain. Furthermore, considering the limitations inherent in existing publication bias detection methods, it is recommended to leverage advancements in clinical evidence research and construct a model that can more effectively detect and correct publication bias. However, it is worth noting that presently there is no AI/ML-based method available for this purpose**.**

IX. CONCLUSIONS AND PRACTICAL IMPLICATIONS

This thesis examines six hypotheses concerning the systematic reviews and evidence syntheses of digital biomarkers in comparison to non-digital biomarkers or pharmaceuticals. Hypothesis I focuses on the classification system and reveals that while the World Health Organization (WHO) tools, including ICD-11, ICHI, and ICF, are not suitable for categorizing population, intervention, outcome, and digital biomarkers in studies, they can effectively be used in studies involving individuals with specific clinical conditions. However, are not applicable for categorizing studies including healthy populations.

Hypothesis II investigates the power of direct and indirect digital biomarker studies and concludes that there is no significant difference. Furthermore, the analysis demonstrates that a majority of studies lack sufficient statistical power.

Hypotheses III and IV address the methodological quality and quality of evidence in systematic reviews of DBMs compared to non-DBMs or pharmaceuticals. The findings indicate that there is no significant difference in terms of methodological quality and quality of evidence between these two groups. Additionally, both groups exhibit low methodological quality and quality of evidence.

Hypotheses V and VI focus on comparing MAs of DBMs with those of non-DBMs or pharmaceuticals. The results suggest that there is no significant difference in terms of publication bias between these groups.

IX.1 Classification of DBM studies using WHO tools

The field of digital health is characterized by rapid development, driven by the continuous evolution of technology. Within this realm, it is noteworthy that diverse technological solutions may coexist to address a given health problem, while a singular solution may be customized to suit specific subgroups. Ascertaining the optimal digital intervention poses a challenge for medical and financial decision-makers, surpassing the complexities encountered in familiarizing oneself with thousands of drugs. To tackle clinical queries and facilitate decision-making, the formulation of PICO (Patient, Intervention, Comparison, Outcome) statements aids in structuring clinical questions (Huang, Lin and Demner-Fushman, 2006). Our observation on the coverage of populations by WHO systems, revealing an emphasis on populations with illness while offering limited coverage for healthy populations.

IX.1.1 Healthy populations are not covered

The prevention of disease during periods of population health is a significant objective pursued by governmental bodies, health systems, and related stakeholders (Haslam, 2014). Maintaining individuals in a state of wellness is likely to be a more financially prudent approach than providing treatment after the onset of illness (Adepoju, Preston and Gonzales, 2015). The utilization of traditional medical technologies primarily occurs in response to illness, whereas the advent of digital technologies has opened up vast possibilities for preventive interventions, which is known as the evolution from treatment to prevention by digital healthcare (Park *et al.*, 2019). The World Health Organization (WHO) suggests that the adoption of digital health technologies, such as wearables, can facilitate lifestyle modifications aimed at promoting preventive measures (N. Khan *et al.*, 2017). This is a huge and unique market opportunity for digital developers, corresponding to the modern definition of health by WHO, as health encompasses *"a comprehensive state of physical, mental, and social well-being, extending beyond the mere absence of disease or infirmity"*.

Nonetheless, the absence of official classification codes for categorizing digital interventions targeting healthy or at-risk populations presents challenges in terms of their adoption, utilization, and financing within existing healthcare systems. Governments, health systems, and other relevant stakeholders should acknowledge and appreciate the substantial potential of digital tools in disease prevention, and accordingly modify their administrative frameworks to effectively incorporate and harness the capabilities of these tools.

IX.1.2 Sick populations are well covered

The rapidly evolving technological landscape, coupled with the proliferation of diverse brand names (e.g., Fitbit, Garmin, Jawbone), presents challenges in market characterization, identification, and the development of new digital tools. Furthermore, clinicians and researchers may encounter difficulties in comprehending the breadth of evidence pertaining to these technologies (Cuff, 2023). Therefore, the establishment of a comprehensive coding system could significantly facilitate the identification and analysis of clinical evidence. The consistent utilization of World Health Organization (WHO) tools in coding Patient, Intervention, Comparison, and Outcome (PICO) statements would streamline the execution of systematic literature reviews (SLRs) and evidence syntheses. While the coding of clinical outcomes was less onerous with a limited number of drugs, the advent of digital transformation necessitates the coding of clinical results to ensure efficient evaluation and integration of digital interventions into healthcare practice.

IX.2 Statistical power of direct – indirect biomarkers

IX.2.1 No difference in power of direct and indirect DBM stusdies

Regulatory and reimbursement agencies commonly distinguish digital technologies and medical devices based on their level of risk, a differentiation that impacts the authorization and reimbursement processes. Devices categorized as high risk, known as direct DBMs, are those that execute interventions without requiring human interaction. On the other hand, devices involving human interaction are considered lower risk, termed indirect DBMs. The authorization and reimbursement of high-risk devices necessitate a greater volume of clinical evidence (Zah *et al.*, 2022; NICE, 2023). These requirements encompass key aspects, including sample size planning, which are integral to the regulatory framework. Consequently, one would anticipate better planned studies when dealing with direct DBMs. The absence of discernible disparities in the statistical power of clinical studies across technologies that potentially belong to disparate risk groups raises concerns regarding the efficacy of various regulatory policies. In fact, the implementation of the new European Medical Device Regulation (MDR) has recently been delayed due to firms struggling to meet the elevated regulatory standards (Mezher, 2023). The proposed MDR aims to enhance patient safety and ensure improved quality standards, emphasizing the need for firms to invest additional efforts in meeting these standards, particularly when developing high-risk devices.

IX.2.2 Low powered studies in both direct and indirect DBMs

Inadequate statistical power poses a significant concern as it leads to resource wastage in the context of digital intervention development. Insufficient power undermines the efficacy evaluation of interventions, devastating the considerable investments made in technical and clinical development, as well as regulatory efforts, vulnerable to loss if negative trial outcomes emerge (Ellis, 2010). While digital developers, often smaller firms compared to pharmaceutical counterparts, may encounter limitations in conducting well-powered trials due to resource constraints, this jeopardizes the viability of their investments. Investors, owners, CEOs, and other stakeholders should consider this issue, emphasizing the imperative for robust clinical studies.

The overestimation of the effect size of an intervention contributes to the issue of underpowered studies. Particularly in the case of indirect DBMs, where low effect sizes are expected, accounting for the complex human factors becomes challenging. This raises questions about the extent to which developers of direct DBMs possess a comprehensive understanding of their technology and the ability to appropriately plan effect sizes for their

studies. Consequently, an overall enhancement in expertise pertaining to the development and execution of clinical studies within the device industry is warranted. Investors, in turn, can utilize the statistical power of clinical studies as an indicator of potential returns on investment, highlighting the importance of robust study designs and sample sizes.

Low statistical power can also introduce publication bias, commonly known as the "small study effect," which can distort the findings of meta-analyses by favoring studies with larger effect sizes. If selective publication occurs, only small studies with substantial effect sizes are likely to be published, potentially inflating the overall effect observed in a meta-analysis. Consequently, this bias can contribute to overestimated expectations regarding the efficacy of medical technologies. To counteract this, the collection of real-world evidence assumes particular significance, prompting a shift towards the improved utilization of real-world data even among regulatory agencies. Companies thus should design their devices and systems to facilitate the collection of real-world data and foster integration within digital health data ecosystems, including considerations of interoperability. This emerging trend represents a significant development in the field.

IX.3 Digital vs non-digital: quality or SLR methods, quality of evidence, publication bias

IX.3.1 No difference between the two groups

Irrespective of variations in industry structure, technology, and regulatory standards (Iqbal and Biller-Andorno, 2022), stakeholders such as clinicians, payers, and patients hold a common expectation for robust supporting evidence when it comes to the utilization of technologies within healthcare systems (Greaves *et al.*, 2018). Developers of digital technologies must recognize and acknowledge this fundamental requirement, as they will eventually face the necessity of meeting evidence standards akin to those imposed on pharmaceuticals. This realization poses a considerable management and investment challenge for smaller, more innovative firms, highlighting the need for careful strategic planning and resource allocation.

IX.3.2 Low overall quality

The matter is of lesser significance when considering the reporting quality of systematic literature reviews (SLRs), as it primarily depends on the expertise and capabilities of researchers rather than the nature of the evidence itself. However, concerns arise regarding the adequacy of the available evidence due to the presence of non-high GRADE studies and a notable number of missing studies, which can potentially introduce publication bias. Moreover, the diminishing productivity of pharmaceutical innovation exacerbates the already high and escalating costs associated with novel advancements. While the attainment of better evidence

appears to correlate with increased costs, the absence of sufficient evidence may impede the adoption of technologies in treatment guidelines and public financing initiatives, ultimately hindering the widespread implementation of innovative solutions. Hence, the generation of non-high quality evidence may ultimately contribute to a wasteful utilization of resources during the development process.

The presence of numerous missing studies serves as an indicator of potential publication bias, a notion further supported by the substantial prevalence of underpowered studies. Although a marginal percentage shift in overall effect sizes may not appear substantial or statistically significant at first glance, its implications can be amplified when integrated into health economic evaluations and models. Ultimately, even a relatively modest level of publication bias can contribute to an inefficient allocation of resources, resulting in wastage within public expenditure. Such misallocation has the potential to undermine public welfare and compromise the optimal utilization of available resources.

IX.4 Conclusion

In recent years, the regulatory standards for medical devices in Europe, specifically through the introduction of the Medical Device Regulation (MDR), have undergone substantial enhancements aimed at addressing the aforementioned concerns and elevating the quality of clinical evidence associated with medical devices. Despite encountering significant implementation challenges and delays surrounding the MDR, the requirements for evidence within the realm of digital health technologies are progressively escalating. Companies that successfully adhere to these heightened clinical standards will not only endure but also gain access to expansive interconnected markets, while those unable to meet these standards may risk losing their market authorization. Therefore, the ability to strategically plan and execute comprehensive clinical studies to generate high-quality clinical evidence is likely to arise as a key source of competitive advantage for firms operating in the medical technology industry. The development of these indispensable abilities requires immediate attention and investment. Investors active in the digital health sector should closely monitor the quality of evidence and clinical trial competencies exhibited by industry participants, as these factors can considerably contribute to the sector's overall business risk.

X. MY OWN PUBLICATIONS

X.1 Journal articles

X.1.1Articles

1. Motahari-Nezhad, Hossein. 2023. “An Artificial Neural Network (ANN) Model for Publication Bias: A Machine Learning-Based Study on PubMed Meta-Analyses.” Aslib Journal of Information Management. doi: 10.1108/AJIM-08-2022-0364.

2. Motahari-Nezhad, Hossein, Hana Al-Abdulkarim, Meriem Fgaier, Mohamed Mahdi Abid, Márta Péntek, László Gulácsi, and Zsombor Zrubka. 2022. “Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews.” Journal of Medical Internet Research 24(12):e41042. doi: 10.2196/41042.

3. Motahari-nezhad, Hossein, Mostafa Miribonjar, and Aslan Sadeghdaghighi. 2022. “Methodological and Evidence Synthesis Quality Evaluation of Meta- Analyses Assessing the Effect of Antibacterial Envelopes to Reduce CIED- Related Infections.” International Journal of Medical Research & Health Sciences 11(2):65–76.

4. Motahari-Nezhad, Hossein, Márta Péntek, László Gulácsi, and Zsombor Zrubka. 2021. “Outcomes of Digital Biomarker–Based Interventions: Protocol for a Systematic Review of Systematic Reviews.” JMIR Research Protocols 10(11):e28204. doi: 10.2196/28204.

X.1.2 Survey paper

1. Motahari-Nezhad, Hossein, and Aslan Sadeghdaghighi. 2023. “Publication Bias in Meta-Analyses of the Therapeutic Efficacy of Remdesivir Interventions for Patients with COVID-19.” Global Knowledge, Memory and Communication. doi: 10.1108/gkmc-02-2022-0030.

2. Motahari-Nezhad, Hossein, Fateme Zare, Hadi Akbari, and Aslan Sadeghdaghighi. 2022. “Health Outcomes of Fitbit, Garmin or Apple Watch-Based Interventions: A Systematic Review of Systematic Reviews.” Baltic Journal of Health and Physical Activity.

3. Motahari-Nezhad, Hossein, Meriem Fgaier, Mohamed Mahdi Abid, Márta Péntek, László Gulácsi, and Zsombor Zrubka. 2022. “Digital Biomarker-Based Studies: Scoping Review of Systematic Reviews.” JMIR MHealth and UHealth 10(10). doi: 10.2196/35722.

X.1.3 Abstract

1. Motahari-Nezhad, H, M. Fgaier, M. Péntek, L. Gulácsi, and Z. Zrubka. 2022. “MT14 Populations, Interventions, and Outcomes in Digital Biomarker-Based Interventions’ Systematic Reviews: A Scoping Review.” Value in Health 25(7):S534. doi: 10.1016/j.jval.2022.04.1293.

2. Mahdi Abid, M., H. Motahari-Nezhad, L. Gulácsi, M. Péntek, and Z. Zrubka. 2022. “POSC36 Implantable Cardiac Defibrillators in Heart Failure Patients: An Overview of Reviews and Meta-Analysis.” Value in Health 25(1):S39. doi: 10.1016/j.jval.2021.11.182.

3. Zrubka, Zsombor, Anita Burrell, Menna N. Sharkawy, Colin M. Pfeiffer, Manthan D. Janodia, Mete Saylan, Cecile van Steen, Hossein Motahari-Nezhad, and Ken Redekop. 2021. “PP424 Piloting A Comprehensive Search For EHealth Definitions In The Grey Literature: Preliminary Results From A Systematic Scoping Review.” International Journal of Technology Assessment in Health Care 37(S1):36–36. doi: 10.1017/S0266462321001562.

X.2 Chapter in book

X.2.1 Conference paper

1. Motahari Nezhad, Hossein, and Zsombor Zrubka. 2022. “Quality of Evidence Assessment of Direct and Indirect Digital Biomarkers” edited by Á. Csiszárik-Kocsir, A. Popovics, and P. Fehér-Polgár. XVII. FIKUSZ 2022 International Conference 663-670 PG – 8.

XI. References

Aarts, S., van den Akker, M. and Winkens, B. (2014) 'The importance of effect sizes', *European Journal of General Practice*, 20(1), pp. 61–64. doi: 10.3109/13814788.2013.818655.

Abdel Shaheed, C. *et al.* (2021) 'The efficacy and safety of paracetamol for pain relief: an overview of systematic reviews', *Medical Journal of Australia*, 214(7), pp. 324–331. doi: 10.5694/mja2.50992.

Abdollahi, S. *et al.* (2020) 'Zinc Supplementation and Body Weight: A Systematic Review and Dose-Response Meta-analysis of Randomized Controlled Trials', *Advances in Nutrition*, 11(2), pp. 398–411. doi: 10.1093/advances/nmz084.

Abdulla, J., Køber, L., *et al.* (2006) 'Effect of beta-blocker therapy on functional status in patients with heart failure - A meta-analysis', *European Journal of Heart Failure*, 8(5), pp. 522–531. doi: 10.1016/j.ejheart.2005.10.012.

Abdulla, J., Haarbo, J., *et al.* (2006) 'Impact of implantable defibrillators and resynchronization therapy on outcome in patients with left ventricular dysfunction - A meta-analysis', *Cardiology*, 106(4), pp. 249–255. doi: 10.1159/000093234.

Abraham, W. T. *et al.* (2016) 'Sustained efficacy of pulmonary artery pressure to guide adjustment of chronic heart failure therapy: complete follow-up results from the CHAMPION randomised trial', *The Lancet*, 387(10017), pp. 453–461. doi: 10.1016/S0140-6736(15)00723-0.

Adepoju, O. E., Preston, M. A. and Gonzales, G. (2015) 'Health Care Disparities in the Post–Affordable Care Act Era', *American Journal of Public Health*, 105(S5), pp. S665–S667. doi: 10.2105/AJPH.2015.302611.

Agrawal, S. *et al.* (2016) 'The role of implantable cardioverter-defibrillators in patients with continuous flow left ventricular assist devices — A meta-analysis', *International Journal of Cardiology*, 222, pp. 379–384. doi: 10.1016/j.ijcard.2016.07.257.

Akel, T. and Lafferty, J. (2017) 'Implantable cardioverter defibrillators for primary prevention in patients with nonischemic cardiomyopathy: A systematic review and meta-analysis', *Cardiovascular Therapeutics*, 35(3), pp. 1–8. doi: 10.1111/1755-5922.12253.

Al-Khatib, S. M. *et al.* (2017) 'Primary Prevention Implantable Cardioverter Defibrillators in Patients With Nonischemic Cardiomyopathy', *JAMA Cardiology*, 2(6), p. 685. doi: 10.1001/jamacardio.2017.0630.

Alba, A. C. *et al.* (2013) 'Predictors of Mortality in Patients With an Implantable Cardiac Defibrillator: A Systematic Review and Meta-analysis', *Canadian Journal of Cardiology*, 29(12), pp. 1729–1740. doi: 10.1016/j.cjca.2013.09.024.

Alba, A. C., Foroutan, F., *et al.* (2018) 'Implantable cardiac defibrillator and mortality in non-ischaemic cardiomyopathy: An updated meta-analysis', *Heart*, 104(3), pp. 230–236. doi: 10.1136/heartjnl-2017-311430.

Alba, A. C., Walter, S. D., *et al.* (2018) 'Predicting Survival in Patients With Heart Failure With an Implantable Cardioverter Defibrillator: The Heart Failure Meta-Score', *Journal of*

Cardiac Failure, 24(11), pp. 735–745. doi: 10.1016/j.cardfail.2017.11.002.

Alexandridis, G. *et al.* (2016) 'An integrated workflow of biomimetic design, material selection and computer aided engineering', *Academic Journal of Manufacturing Engineering*, 14(4), pp. 12–18.

Aljunid, S. M. *et al.* (2015) 'ICHI Categorial Structure: A WHO-FIC Tool for Semantic Interoperability of Procedures Classifications', *Studies in Health Technology and Informatics*, 216, p. 1090. doi: 10.3233/978-1-61499-564-7-1090.

Almalik, O., Zhan, Z. and van den Heuvel, E. R. (2021) 'Tests for publication bias are unreliable in case of heteroscedasticity', *Contemporary Clinical Trials Communications*, 22(April), p. 100781. doi: 10.1016/j.conctc.2021.100781.

Alotaibi, S. *et al.* (2020) 'Remote monitoring of implantable cardiac devices in heart failure patients: a systematic review and meta-analysis of randomized controlled trials', *Heart Failure Reviews*, 25(3), pp. 469–479. doi: 10.1007/s10741-020-09923-1.

Alskaf, E. E., Tridente, A. A. and Al-Mohammad, A. A. (2016) 'Tolvaptan for Heart Failure, Systematic Review and Meta-Analysis of Trials', *Journal of Cardiovascular Pharmacology*, 68(3), pp. 196–203. doi: 10.1097/FJC.0000000000000405.

AlTurki, A. *et al.* (2019) 'Implantable cardioverter-defibrillator use in elderly patients receiving cardiac resynchronization: A meta-analysis', *Hellenic Journal of Cardiology*, 60(5), pp. 276–281. doi: 10.1016/j.hjc.2017.12.003.

An, H. R. *et al.* (2022) 'Meta-Analysis of the Factors Influencing the Restoration of Spontaneous Circulation After Cardiopulmonary Resuscitation', *Frontiers in Physiology*, 13(March). doi: 10.3389/fphys.2022.834352.

Anantha Narayanan, M. *et al.* (2017) 'Efficacy of Implantable Cardioverter-Defibrillator Therapy in Patients With Nonischemic Cardiomyopathy', *JACC: Clinical Electrophysiology*, 3(9), pp. 962–970. doi: 10.1016/j.jacep.2017.02.006.

Ang, S. M. *et al.* (2021) 'Efficacy of Interventions That Incorporate Mobile Apps in Facilitating Weight Loss and Health Behavior Change in the Asian Population: Systematic Review and Meta-analysis', *Journal of Medical Internet Research*, 23(11), p. e28185. doi: 10.2196/28185.

Anker, S. D. *et al.* (2018) 'Effects of ferric carboxymaltose on hospitalisations and mortality rates in iron-deficient heart failure patients: an individual patient data meta-analysis', *European Journal of Heart Failure*, 20(1), pp. 125–133. doi: 10.1002/ejhf.823.

Antoun, J. *et al.* (2022) 'The Effectiveness of Combining Nonmobile Interventions With the Use of Smartphone Apps With Various Features for Weight Loss: Systematic Review and Meta-analysis', *JMIR mHealth and uHealth*, 10(4), pp. 1–22. doi: 10.2196/35479.

Arastu, N. *et al.* (2022) 'Efficacy of subcutaneous semaglutide compared to placebo for weight loss in obese, non-diabetic adults: a systematic review & meta-analysis', *International Journal of Clinical Pharmacy*, 44(4), pp. 852–859. doi: 10.1007/s11096-022-01428-1.

Armstrong, M. *et al.* (2019) 'Use of pedometers as a tool to promote daily physical activity levels in patients with COPD: a systematic review and meta-analysis', *European Respiratory Review*, 28(154), p. 190039. doi: 10.1183/16000617.0039-2019.

Arundale, A. J. H. *et al.* (2018) 'Exercise-based knee and anterior cruciate ligament injury prevention', *Journal of Orthopaedic and Sports Physical Therapy*, 48(9), pp. A1–A25. doi: 10.2519/jospt.2018.0303.

Avgerinos, I. *et al.* (2020) 'Oral semaglutide for type 2 diabetes: A systematic review and meta-

analysis', *Diabetes, Obesity and Metabolism*, 22(3), pp. 335–345. doi: 10.1111/dom.13899.

Babrak, L. M. *et al.* (2019) 'Traditional and Digital Biomarkers: Two Worlds Apart?', *Digital Biomarkers*, 3(2), pp. 92–102. doi: 10.1159/000502000.

Bamforth, R. J. *et al.* (2021) 'Strategies to prevent hospital readmission and death in patients with chronic heart failure, chronic obstructive pulmonary disease, and chronic kidney disease: A systematic review and meta-analysis', *PLoS ONE*, 16(4 April), pp. 1–26. doi: 10.1371/journal.pone.0249542.

Barakat, A. F. *et al.* (2017) 'Primary prevention implantable cardioverter defibrillator in patients with non-ischaemic cardiomyopathy: A meta-analysis of randomised controlled trials', *BMJ Open*, 7(6), pp. 1–8. doi: 10.1136/bmjopen-2017-016352.

Barch, D. M. and Yarkoni, T. (2013) 'Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research', *Cognitive, Affective, & Behavioral Neuroscience*, 13(4), pp. 687–689. doi: 10.3758/s13415-013-0201-7.

Barra, S. *et al.* (2015) 'Importance of Implantable Cardioverter-Defibrillator Back-Up in Cardiac Resynchronization Therapy Recipients: A Systematic Review and Meta-Analysis', *Journal of the American Heart Association*, 4(11). doi: 10.1161/JAHA.115.002539.

Barra, S. *et al.* (2018) 'Cause-of-death analysis in patients with cardiac resynchronization therapy with or without a defibrillator: a systematic review and proportional meta-analysis', *EP Europace*, 20(3), pp. 481–491. doi: 10.1093/europace/eux094.

Baskerville, R. *et al.* (2017) 'Impact of accelerometer and pedometer use on physical activity and glycaemic control in people with Type 2 diabetes: a systematic review and meta-analysis', *Diabetic Medicine*, 34(5), pp. 612–620. doi: 10.1111/dme.13331.

Bastian, H., Glasziou, P. and Chalmers, I. (2010a) 'Seventy-five trials and eleven systematic reviews a day: How will we ever keep up?', *PLoS Medicine*, 7(9). doi: 10.1371/journal.pmed.1000326.

Bastian, H., Glasziou, P. and Chalmers, I. (2010b) 'Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?', *PLoS Medicine*, 7(9), p. e1000326. doi: 10.1371/journal.pmed.1000326.

Batchelor, R. J. *et al.* (2023) 'Meta-Analysis on Drug and Device Therapy of New York Heart Association Functional Class IV Heart Failure With Reduced Ejection Fraction', *The American Journal of Cardiology*, 188, pp. 52–60. doi: 10.1016/j.amjcard.2022.11.001.

Bazoukis, G. *et al.* (2019) 'Impact of Implantable Cardioverter-Defibrillator Interventions on All-Cause Mortality in Heart Failure Patients', *Cardiology in Review*, 27(3), pp. 160–166. doi: 10.1097/CRD.0000000000000226.

Bedaso, A. *et al.* (2021) 'The relationship between social support and mental health problems during pregnancy: a systematic review and meta-analysis', *Reproductive Health*, 18(1), pp. 1–23. doi: 10.1186/s12978-021-01209-5.

Berbenetz, N. M. and Mrkobrada, M. (2016) 'Mineralocorticoid receptor antagonists for heart failure: Systematic review and meta-analysis', *BMC Cardiovascular Disorders*, 16(1), pp. 1–10. doi: 10.1186/s12872-016-0425-x.

Bergau, L., Seegers, J. and Zabel, M. (2014) 'Sex differences in ICD benefit', *Journal of Electrocardiology*, 47(6), pp. 869–873. doi: 10.1016/j.jelectrocard.2014.08.007.

Bertoldi, E. G. *et al.* (2011) 'Mortality reduction of cardiac resynchronization and implantable cardioverter-defibrillator therapy in heart failure: An updated meta-analysis. does recent

evidence change the standard of care?', *Journal of Cardiac Failure*, 17(10), pp. 860–866. doi: 10.1016/j.cardfail.2011.06.372.

Betts, T. R. *et al.* (2013) 'Absolute risk reduction in total mortality with implantable cardioverter defibrillators: analysis of primary and secondary prevention trial data to aid risk/benefit analysis', *Europace*, 15(6), pp. 813–819. doi: 10.1093/europace/eus427.

Bezeau, S. and Graves, R. (2001) 'Statistical Power and Effect Sizes of Clinical Neuropsychology Research', *Journal of Clinical and Experimental Neuropsychology*, 23(3), pp. 399–406. doi: 10.1076/jcen.23.3.399.1181.

BisResearch (2020) *Global Digital Biomarkers Market Focus on Key Trends, Growth Potential, Competitive Landscape, Components (Data Collection and Integration), End Users, Application (Sleep and Movement, Neuro, Respiratory and Cardiological Disorders) and Region – Analysis* . Available at: https://bisresearch.com/industry-report/digital-biomarkers-market.html (Accessed: 27 December 2020).

Blough, J. and Loprinzi, P. D. (2018) 'Experimentally investigating the joint effects of physical activity and sedentary behavior on depression and anxiety: A randomized controlled trial', *Journal of Affective Disorders*, 239(May), pp. 258–268. doi: 10.1016/j.jad.2018.07.019.

Bonnechère, B. *et al.* (2021) 'Is mhealth a useful tool for self-assessment and rehabilitation of people with multiple sclerosis? A systematic review', *Brain Sciences*, 11(9). doi: 10.3390/brainsci11091187.

Borde, R. *et al.* (2017) 'Methodological considerations and impact of school-based interventions on objectively measured physical activity in adolescents: a systematic review and meta-analysis', *Obesity Reviews*, 18(4), pp. 476–490. doi: 10.1111/obr.12517.

Bourke, M. *et al.* (2023) 'Adherence to the World Health Organization's physical activity recommendation in preschool-aged children: a systematic review and meta-analysis of accelerometer studies', *International Journal of Behavioral Nutrition and Physical Activity*, 20(1), p. 52. doi: 10.1186/s12966-023-01450-0.

Braakhuis, H., Berger, M. and Bussmann, J. (2019) 'Effectiveness of healthcare interventions using objective feedback on physical activity: A systematic review and meta-analysis', *Journal of Rehabilitation Medicine*, 51(3), pp. 151–159. doi: 10.2340/16501977-2522.

Bradley, D. J. *et al.* (2003) 'Cardiac Resynchronization and Death From Progressive Heart Failure', *JAMA*, 289(6), p. 730. doi: 10.1001/jama.289.6.730.

Brok, J. *et al.* (2008) 'Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses', *Journal of Clinical Epidemiology*, 61(8), pp. 763–769. doi: 10.1016/j.jclinepi.2007.10.007.

Brok, J. *et al.* (2009) 'Apparently conclusive meta-analyses may be inconclusive—Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses', *International Journal of Epidemiology*, 38(1), pp. 287–298. doi: 10.1093/ije/dyn188.

Bussadori, S. K. *et al.* (2020) 'The current trend in management of bruxism and chronic pain: An overview of systematic reviews', *Journal of Pain Research*, 13, pp. 2413–2421. doi: 10.2147/JPR.S268114.

Butler, J. *et al.* (2020) 'Efficacy and safety of SGLT2 inhibitors in heart failure: systematic review and meta-analysis', *ESC Heart Failure*, 7(6), pp. 3298–3309. doi: 10.1002/ehf2.13169.

Button, K. S. *et al.* (2013) 'Power failure: why small sample size undermines the reliability of

neuroscience', *Nature Reviews Neuroscience*, 14(5), pp. 365–376. doi: 10.1038/nrn3475.

Cai, X. *et al.* (2016) 'Pedometer intervention and weight loss in overweight and obese adults with Type 2 diabetes: a meta-analysis', *Diabetic Medicine*, 33(8), pp. 1035–1044. doi: 10.1111/dme.13104.

Cai, X. *et al.* (2020) 'Mobile Application Interventions and Weight Loss in Type 2 Diabetes: A Meta-Analysis', *Obesity*, 28(3), pp. 502–509. doi: 10.1002/oby.22715.

Campo, G. *et al.* (2017) 'Clinical benefit of drugs targeting mitochondrial function as an adjunct to reperfusion in ST-segment elevation myocardial infarction: A meta-analysis of randomized clinical trials', *International Journal of Cardiology*, 244, pp. 59–66. doi: 10.1016/j.ijcard.2017.06.040.

Carmo, A. A. L. *et al.* (2018) 'Implantable cardioverter-defibrillator in Chagas heart disease: A systematic review and meta-analysis of observational studies', *International Journal of Cardiology*, 267, pp. 88–93. doi: 10.1016/j.ijcard.2018.05.091.

Chan, K. L. and Chen, M. (2019) 'Effects of Social Media and Mobile Health Apps on Pregnancy Care: Meta-Analysis', *JMIR mHealth and uHealth*, 7(1), p. e11836. doi: 10.2196/11836.

CHATTERJEE, S. *et al.* (2013) 'Pharmacologic Rate versus Rhythm-Control Strategies in Atrial Fibrillation: An Updated Comprehensive Review and Meta-Analysis', *Pacing and Clinical Electrophysiology*, 36(1), pp. 122–133. doi: 10.1111/j.1540-8159.2012.03513.x.

Chen, Shaojie *et al.* (2013) 'The efficacy and safety of cardiac resynchronization therapy combined with implantable cardioverter defibrillator for heart failure: a meta-analysis of 5674 patients', *Europace*, 15(7), pp. 992–1001. doi: 10.1093/europace/eus419.

Chen, S. *et al.* (2013) 'The efficacy and safety of cardiac resynchronization therapy combined with implantable cardioverter defibrillator for heart failure: a meta-analysis of 5674 patients', *Europace*, 15(7), pp. 992–1001. doi: 10.1093/europace/eus419.

Chew, H. S. J. *et al.* (2022) 'Sustainability of Weight Loss Through Smartphone Apps: Systematic Review and Meta-analysis on Anthropometric, Metabolic, and Dietary Outcomes', *Journal of Medical Internet Research*, 24(9). doi: 10.2196/40141.

Chew, H. S. J. *et al.* (2023) 'Effectiveness of Combined Health Coaching and Self-Monitoring Apps on Weight-Related Outcomes in People With Overweight and Obesity: Systematic Review and Meta-analysis', *Journal of Medical Internet Research*, 25, p. e42432. doi: 10.2196/42432.

Chiavaroli, L. *et al.* (2019) 'DASH dietary pattern and cardiometabolic outcomes: An umbrella review of systematic reviews and meta-analyses', *Nutrients*, 11(2). doi: 10.3390/nu11020338.

Cleland, J. G. *et al.* (2013) 'An individual patient meta-analysis of five randomized trials assessing the effects of cardiac resynchronization therapy on morbidity and mortality in patients with symptomatic heart failure', *European Heart Journal*, 34(46), pp. 3547–3556. doi: 10.1093/eurheartj/eht290.

Cleland, J. G. F. *et al.* (2022) 'The effect of cardiac resynchronization without a defibrillator on morbidity and mortality: an <scp>individual patient data meta-analysis</scp> of <scp>COMPANION</scp> and <scp>CARE-HF</scp>', *European Journal of Heart Failure*, 24(6), pp. 1080–1090. doi: 10.1002/ejhf.2524.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, *Statistical Power Analysis for the Behavioral Sciences*. Routledge. doi: 10.4324/9780203771587.

Cohn, L. D. and Becker, B. J. (2003) 'How meta-analysis increases statistical power.', *Psychological Methods*, 8(3), pp. 243–253. doi: 10.1037/1082-989X.8.3.243.

Coravos, A., Khozin, S. and Mandl, K. D. (2019) 'Developing and adopting safe and effective digital biomarkers to improve patient outcomes', *npj Digital Medicine*, 2(1), p. 14. doi: 10.1038/s41746-019-0090-4.

Cortés, O. L. *et al.* (2021) 'Systematic review and meta-analysis of clinical trials: In-hospital use of sensors for prevention of falls', *Medicine*, 100(41), p. e27467. doi: 10.1097/MD.0000000000027467.

Croteau, K. A. *et al.* (2007) 'Effect of a Pedometer-Based Intervention on Daily Step Counts of Community-Dwelling Older Adults', *Research Quarterly for Exercise and Sport*, 78(5), pp. 401–406. doi: 10.1080/02701367.2007.10599439.

Cuff, A. (2023) 'The evolution of digital health and its continuing challenges', *BMC Digital Health*, 1(1), p. 3. doi: 10.1186/s44247-022-00004-x.

Cui, M. *et al.* (2016) 'T2DM self-management via smartphone applications: A systematic review and meta-analysis', *PLoS ONE*, 11(11), pp. 1–15. doi: 10.1371/journal.pone.0166718.

Daubert, J.-C., Martins, R. and Leclercq, C. (2015) 'Why We Have to Use Cardiac Resynchronization Therapy–Pacemaker More', *Cardiac Electrophysiology Clinics*, 7(4), pp. 709–720. doi: 10.1016/j.ccep.2015.08.016.

Dehghan Ghahfarokhi, A. *et al.* (2022) 'The Effect of Wearable and Smartphone Applications on Physical Activity, Quality of Life, and Cardiovascular Health Outcomes in Overweight/Obese Adults: A Systematic Review and Meta-analysis of Randomized Controlled Trials', *Biological Research for Nursing*, 24(4), pp. 503–518. doi: 10.1177/10998004221099556.

DeVito, N. J. and Goldacre, B. (2019) 'Catalogue of bias: publication bias', *BMJ Evidence-Based Medicine*, 24(2), pp. 53–54. doi: 10.1136/bmjebm-2018-111107.

DiFilippo, K. N. *et al.* (2015) 'The use of mobile apps to improve nutrition outcomes: A systematic literature review', *Journal of Telemedicine and Telecare*, 21(5), pp. 243–253. doi: 10.1177/1357633X15572203.

Dinicolantonio, J. J. *et al.* (2013) 'Meta-analysis of carvedilol versus beta 1 selective beta-blockers (atenolol, bisoprolol, metoprolol, and nebivolol)', *American Journal of Cardiology*, 111(5), pp. 765–769. doi: 10.1016/j.amjcard.2012.11.031.

drugbank (2022) *Drug Statistics*. Available at: https://go.drugbank.com/stats (Accessed: 20 August 2022).

Du, S. *et al.* (2020) 'The efficacy of e-health in the self-management of chronic low back pain: A meta analysis', *International Journal of Nursing Studies*, 106, p. 103507. doi: 10.1016/j.ijnurstu.2019.103507.

Duan, L. *et al.* (2020) 'Effects of Vitamin D Supplementation on General and Central Obesity: Results from 20 Randomized Controlled Trials Involving Apparently Healthy Populations', *Annals of Nutrition and Metabolism*, 76(3), pp. 153–164. doi: 10.1159/000507418.

Duffy, A., Christie, G. J. and Moreno, S. (2022) 'The Challenges Toward Real-world Implementation of Digital Health Design Approaches: Narrative Review', *JMIR Human Factors*, 9(3), p. e35693. doi: 10.2196/35693.

Dumit, E. M. *et al.* (2018) 'The use of eHealth with immunizations: An overview of systematic reviews', *Vaccine*, 36(52), pp. 7923–7928. doi: 10.1016/j.vaccine.2018.06.076.

Dunn, R. E., White, C. R. and Green, J. A. (2018) 'A model to estimate seabird field metabolic rates', *Biology Letters*, 14(6), p. 20180190. doi: 10.1098/rsbl.2018.0190.

Duval, S. and Tweedie, R. (2000) 'Trim and Fill: A Simple Funnel-Plot-Based Method', *Biometrics*, 56(June), pp. 455–463.

Earley, A. *et al.* (2014) 'Effectiveness of Implantable Cardioverter Defibrillators for Primary Prevention of Sudden Cardiac Death in Subgroups', *Annals of Internal Medicine*, 160(2), pp. 111–121. doi: 10.7326/M13-1787.

Ebrahim Valojerdi, A., Tanha, K. and Janani, L. (2017) 'Important considerations in calculating and reporting of sample size in randomized controlled trials', *Medical Journal of the Islamic Republic of Iran*, 31(1), pp. 843–846. doi: 10.14196/mjiri.31.127.

Eldawlatly, A. *et al.* (2018) 'Appearance of Population, Intervention, Comparison, and Outcome as research question in the title of articles of three different anesthesia journals: A pilot study', *Saudi Journal of Anaesthesia*, 12(2), pp. 283–286. doi: 10.4103/sja.SJA_767_17.

Elkaryoni, A. *et al.* (2019) 'Implantable cardioverter-defibrillators and survival in advanced heart failure patients with continuous-flow left ventricular assist devices: a systematic review and meta-analysis', *EP Europace*, 21(9), pp. 1353–1359. doi: 10.1093/europace/euz125.

Ellis, P. D. (2010) *The Essential Guide to Effect Sizes*. Cambridge University Press. doi: 10.1017/CBO9780511761676.

EMA (2020) *Human medicines: regulatory information*. Available at: https://www.ema.europa.eu/en/human-medicines-regulatory-information (Accessed: 22 March 2023).

Engstrom, N. *et al.* (2022) 'Fragmented QRS is associated with ventricular arrhythmias in heart failure patients: A systematic review and meta-analysis', *Annals of Noninvasive Electrocardiology*, 27(1). doi: 10.1111/anec.12910.

Eysenbach, G. (2001) 'What is e-health?', *Journal of Medical Internet Research*, 3(2), p. e20. doi: 10.2196/jmir.3.2.e20.

Fakih El Khoury, C. *et al.* (2019) 'The Effects of Dietary Mobile Apps on Nutritional Outcomes in Adults with Chronic Diseases: A Systematic Review and Meta-Analysis', *Journal of the Academy of Nutrition and Dietetics*, 119(4), pp. 626–651. doi: 10.1016/j.jand.2018.11.010.

Fänge, A. M. *et al.* (2020) 'Satisfaction with health care interventions among community dwelling people with cognitive disorders and their informal caregivers—a systematic review', *Healthcare (Switzerland)*, 8(3). doi: 10.3390/healthcare8030240.

Farooq, A. *et al.* (2020) 'Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: A systematic review and meta-analysis', *Obesity Reviews*, 21(1). doi: 10.1111/obr.12953.

Farrugia, P. *et al.* (2010) 'Practical tips for surgical research: Research questions, hypotheses and objectives.', *Canadian journal of surgery. Journal canadien de chirurgie*, 53(4), pp. 278–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20646403%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2912019.

FDA (2020) *What is Digital Health?* Available at: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health (Accessed: 22 July 2022).

FDA (2022) *Guidance, Compliance, & Regulatory Information*. Available at:

https://www.fda.gov/drugs/guidance-compliance-regulatory-information (Accessed: 25 March 2023).

FDA (2023) *Drugs*. Available at: https://www.fda.gov/consumers/consumer-updates/drugs (Accessed: 20 February 2023).

Ferguson, T. *et al.* (2022) 'Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses', *The Lancet Digital Health*, 4(8), pp. e615–e626. doi: 10.1016/S2589-7500(22)00111-X.

Ferretti, A., Ronchi, E. and Vayena, E. (2019) 'From principles to practice: benchmarking government guidance on health apps', *The Lancet Digital Health*, 1(2), pp. e55–e57. doi: 10.1016/S2589-7500(19)30027-5.

Finucane, A. M. *et al.* (2021) 'Digital health interventions in palliative care: a systematic meta-review', *npj Digital Medicine*, 4(1). doi: 10.1038/s41746-021-00430-7.

Flores Mateo, G. *et al.* (2015) 'Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis', *Journal of Medical Internet Research*, 17(11), p. e253. doi: 10.2196/jmir.4836.

Frampton, G. *et al.* (2022) 'Principles and framework for assessing the risk of bias for studies included in comparative quantitative environmental systematic reviews', *Environmental Evidence*, 11(1), p. 12. doi: 10.1186/s13750-022-00264-0.

Franssen, W. M. A. *et al.* (2020) 'Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials', *International Journal of Behavioral Nutrition and Physical Activity*, 17(1), p. 57. doi: 10.1186/s12966-020-00955-2.

Friedman, D. J. *et al.* (2017) 'New York Heart Association class and the survival benefit from primary prevention implantable cardioverter defibrillators: A pooled analysis of 4 randomized controlled trials', *American Heart Journal*, 191, pp. 21–29. doi: 10.1016/j.ahj.2017.06.002.

Fuhr, U. and Hellmich, M. (2015) 'Channeling the flood of meta-analyses', *European Journal of Clinical Pharmacology*, 71(6), pp. 645–647. doi: 10.1007/s00228-015-1838-7.

Fung, K. W. *et al.* (2021) 'Evaluation of the International Classification of Health Interventions (ICHI) in the coding of common surgical procedures', *Journal of the American Medical Informatics Association : JAMIA*, 29(1), pp. 43–51. doi: 10.1093/jamia/ocab220.

Gama, F. *et al.* (2020) 'Implantable cardioverter–defibrillators in trials of drug therapy for heart failure: A systematic review and meta-analysis', *Journal of the American Heart Association*, 9(8). doi: 10.1161/JAHA.119.015177.

Gao, D. *et al.* (2011) 'Trimetazidine: A meta-analysis of randomised controlled trials in heart failure', *Heart*, 97(4), pp. 278–286. doi: 10.1136/hrt.2010.208751.

Gaskin, C. J. and Happell, B. (2014) 'Power, effects, confidence, and significance: An investigation of statistical practices in nursing research', *International Journal of Nursing Studies*, 51(5), pp. 795–806. doi: 10.1016/j.ijnurstu.2013.09.014.

Georgarakis, A.-M., Wolf, P. and Riener, R. (2019) 'Simplifying Exosuits: Kinematic Couplings in the Upper Extremity during Daily Living Tasks', in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, pp. 423–428. doi: 10.1109/ICORR.2019.8779401.

Ghanbari, H. (2009) 'Effectiveness of Implantable Cardioverter-Defibrillators for the Primary Prevention of Sudden Cardiac Death in Women With Advanced Heart Failure', *Archives of*

Internal Medicine, 169(16), p. 1500. doi: 10.1001/archinternmed.2009.255.

Gigante, A. *et al.* (2019) 'Acute osteomyelitis and septic arthritis in children: A systematic review of systematic reviews', *European Review for Medical and Pharmacological Sciences*, 23(2), pp. 145–158. doi: 10.26355/EURREV_201904_17484.

Girgis, B. and Duarte, J. A. (2020) 'Physical therapy for tendinopathy: An umbrella review of systematic reviews and meta-analyses', *Physical Therapy in Sport*, 46, pp. 30–46. doi: 10.1016/j.ptsp.2020.08.002.

Goff, D. C. *et al.* (2014) '2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk', *Circulation*, 129(25_suppl_2). doi: 10.1161/01.cir.0000437741.48606.98.

Gold, N. *et al.* (2021) 'Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: Systematic review of reviews', *Journal of Medical Internet Research*, 23(5). doi: 10.2196/19688.

Gopalakrishnan, S. and Ganeshkumar, P. (2013) 'Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare', *Journal of Family Medicine and Primary Care*, 2(1), p. 9. doi: 10.4103/2249-4863.109934.

Greaves, F. *et al.* (2018) 'What is an appropriate level of evidence for a digital health intervention?', *The Lancet*, 392(10165), pp. 2665–2667. doi: 10.1016/S0140-6736(18)33129-5.

Griauzde, D. *et al.* (2019) 'A Mobile Phone-Based Program to Promote Healthy Behaviors Among Adults With Prediabetes Who Declined Participation in Free Diabetes Prevention Programs: Mixed-Methods Pilot Randomized Controlled Trial', *JMIR mHealth and uHealth*, 7(1), p. e11267. doi: 10.2196/11267.

Guo, C. *et al.* (2020) 'Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches', *npj Digital Medicine*, 3(1), pp. 1–14. doi: 10.1038/s41746-020-00314-2.

Guo, X. *et al.* (2022) 'The Antiobesity Effect and Safety of GLP-1 Receptor Agonist in Overweight/Obese Patients Without Diabetes: A Systematic Review and Meta-Analysis', *Hormone and Metabolic Research*, 54(7), pp. 458–471. doi: 10.1055/a-1844-1176.

Guthrie, N. L. *et al.* (2019) 'Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study', *BMJ Open*, 9(7), p. e030710. doi: 10.1136/bmjopen-2019-030710.

Guyatt, G. H. *et al.* (2008) 'GRADE: an emerging consensus on rating quality of evidence and strength of recommendations', *BMJ*, 336(7650), pp. 924–926. doi: 10.1136/bmj.39489.470347.AD.

Guyatt, G. H. (2008) 'Guyatt 2008 - what is "quality of evidence" and why is it important to clinicians.pdf', 336(may), pp. 995–998.

Guyatt, G. H., Oxman, A. D., Vist, G., *et al.* (2011) 'GRADE guidelines: 4. Rating the quality of evidence - Study limitations (risk of bias)', *Journal of Clinical Epidemiology*, 64(4), pp. 407–415. doi: 10.1016/j.jclinepi.2010.07.017.

Guyatt, G. H., Oxman, A. D., Montori, V., *et al.* (2011) 'GRADE guidelines: 5. Rating the quality of evidence - Publication bias', *Journal of Clinical Epidemiology*, 64(12), pp. 1277–1282. doi: 10.1016/j.jclinepi.2011.01.011.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Glasziou, P., *et al.* (2011) 'GRADE guidelines: 7. Rating the quality of evidence -

Inconsistency', *Journal of Clinical Epidemiology*, 64(12), pp. 1294–1302. doi: 10.1016/j.jclinepi.2011.03.017.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Falck-Ytter, Y., *et al.* (2011) 'GRADE guidelines: 8. Rating the quality of evidence - Indirectness', *Journal of Clinical Epidemiology*, 64(12), pp. 1303–1310. doi: 10.1016/j.jclinepi.2011.04.014.

Guyatt, G. H., Oxman, A. D., Kunz, R., Brozek, J., *et al.* (2011) 'GRADE guidelines 6. Rating the quality of evidence - Imprecision', *Journal of Clinical Epidemiology*, 64(12), pp. 1283–1293. doi: 10.1016/j.jclinepi.2011.01.012.

H-Jennings, F. *et al.* (2016) 'Promote Students' Healthy Behavior Through Sensor and Game: A Randomized Controlled Trial', *Medical Science Educator*, 26(3), pp. 349–355. doi: 10.1007/s40670-016-0253-8.

Halawa, A., Enezate, T. and Flaker, G. (2019) 'Device monitoring in heart failure management: outcomes based on a systematic review and meta-analysis', *Cardiovascular Diagnosis and Therapy*, 9(4), pp. 386–393. doi: 10.21037/cdt.2019.01.02.

Hammersley, M. L. *et al.* (2019) 'An internet-based childhood obesity prevention program (TIMe2bhealthy) for parents of preschool-aged children: Randomized controlled trial', *Journal of Medical Internet Research*, 21(2). doi: 10.2196/11964.

Hansen, M. R. *et al.* (2020) 'Postponement of Death by Pharmacological Heart Failure Treatment: A Meta-Analysis of Randomized Clinical Trials', *American Journal of Medicine*, 133(6), pp. e280–e289. doi: 10.1016/j.amjmed.2019.11.015.

Harrison, J. E. *et al.* (2021) 'ICD-11: an international classification of diseases for the twenty-first century', *BMC Medical Informatics and Decision Making*, 21, pp. 1–10. doi: 10.1186/s12911-021-01534-6.

Haslam, D. (2014) 'Prevention is better than cure', *Practitioner*, 258(1769), p. 39.

Head, M. L. *et al.* (2015) 'The Extent and Consequences of P-Hacking in Science', *PLOS Biology*, 13(3), p. e1002106. doi: 10.1371/journal.pbio.1002106.

Her, J. and Cho, M. K. (2021) 'Effect of aromatherapy on sleep quality of adults and elderly people: A systematic literature review and meta-analysis', *Complementary Therapies in Medicine*, 60, p. 102739. doi: 10.1016/j.ctim.2021.102739.

Herring, L. Y. *et al.* (2016) 'Changes in physical activity behaviour and physical function after bariatric surgery: a systematic review and meta-analysis', *Obesity Reviews*, 17(3), pp. 250–261. doi: 10.1111/obr.12361.

Hess, P. L. *et al.* (2013) 'Survival benefit of primary prevention implantable cardioverter-defibrillator therapy after myocardial infarction: Does time to implant matter? A meta-analysis using patient-level data from 4 clinical trials', *Heart Rhythm*, 10(6), pp. 828–835. doi: 10.1016/j.hrthm.2013.02.011.

Higgins, J. P. T. (2003) 'Measuring inconsistency in meta-analyses', *BMJ*, 327(7414), pp. 557–560. doi: 10.1136/bmj.327.7414.557.

Higgins, J. P. T. *et al.* (2019a) *Cochrane handbook for systematic reviews of interventions, Cochrane Handbook for Systematic Reviews of Interventions*. doi: 10.1002/9781119536604.

Higgins, J. P. T. *et al.* (eds) (2019b) *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley. doi: 10.1002/9781119536604.

Higgins, J. P. T., Whitehead, A. and Simmonds, M. (2011) 'Sequential methods for random-effects meta-analysis', *Statistics in Medicine*, 30(9), pp. 903–921. doi: 10.1002/sim.4088.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, W. V. (2022a) *Cochrane Handbook for Systematic Reviews of Interventions*. Available at: www.training.cochrane.org/handbook.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, W. V. (2022b) 'Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022)', in. Chichester (UK).

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, W. V. (2022c) *Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022)*. Available at: www.training.cochrane.org/handbook.

Hindricks, G. *et al.* (2017) 'Daily remote monitoring of implantable cardioverter-defibrillators: insights from the pooled patient-level data from three randomized controlled trials (IN-TIME, ECOST, TRUST)', *European Heart Journal*, 38(22), pp. 1749–1755. doi: 10.1093/eurheartj/ehx015.

Hodkinson, A. *et al.* (2019) 'Accelerometer- and Pedometer-Based Physical Activity Interventions Among Adults With Cardiometabolic Conditions', *JAMA Network Open*, 2(10), p. e1912895. doi: 10.1001/jamanetworkopen.2019.12895.

Hoebert, J. (2015) 'Minds Open – Sustainability of the European regulatory system for medicinal products', (January), pp. 1–15.

Holmen, H. *et al.* (2017) 'Tailored communication within mobile apps for diabetes self-management: A systematic review', *Journal of Medical Internet Research*, 19(6). doi: 10.2196/jmir.7045.

Holmes, D. R. and Wood, D. L. (2006) 'Medical Devices and Pharmaceuticals: Differences in Public Acceptance', *The American Heart Hospital Journal*, 4(4), pp. 269–272. doi: 10.1111/j.1541-9215.2006.06102.x.

Huang, W. L. *et al.* (2018) 'Use of tolvaptan vs. furosemide in older patients with heart failure: Meta-analysis of randomized controlled trials', *Herz*, 43(4), pp. 338–345. doi: 10.1007/s00059-017-4563-4.

Huang, X., Lin, J. and Demner-Fushman, D. (2006) 'Evaluation of PICO as a knowledge representation for clinical questions.', *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2006, pp. 359–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17238363.

Huang, Y. *et al.* (2010) 'All cause mortality of cardiac resynchronization therapy with implantable cardioverter defibrillator: A meta-analysis of randomized controlled trials', *International Journal of Cardiology*, 145(3), pp. 413–417. doi: 10.1016/j.ijcard.2010.05.016.

Ibrahim, M. S. *et al.* (2022) 'Digital health for quality healthcare: A systematic mapping of review studies', *DIGITAL HEALTH*, 8, p. 205520762210858. doi: 10.1177/20552076221085810.

Imberger, G. *et al.* (2015) 'Systematic reviews of anesthesiologic interventions reported as statistically significant: Problems with power, precision, and type 1 error protection', *Anesthesia and Analgesia*, 121(6), pp. 1611–1622. doi: 10.1213/ANE.0000000000000892.

Insel, T. R. (2017) 'Digital Phenotyping', *JAMA*, 318(13), p. 1215. doi: 10.1001/jama.2017.11295.

Ioannidis, J. P. A. (2016) 'The Mass Production of Redundant, Misleading, and Conflicted

Systematic Reviews and Meta-analyses', *Milbank Quarterly*, 94(3), pp. 485–514. doi: 10.1111/1468-0009.12210.

Iqbal, J. D. and Biller-Andorno, N. (2022) 'The regulatory gap in digital health and alternative pathways to bridge it', *Health Policy and Technology*, 11(3), p. 100663. doi: 10.1016/j.hlpt.2022.100663.

IQVIA (2017) *mpact of Digital Health Grows as Innovation, Evidence and Adoption of Mobile Health Apps Accelerate*. Available at: https://www.iqvia.com/newsroom/2017/11/impact-of-digital-health-grows-as-innovation-evidence-and-adoption-of-mobile-health-apps-accelerate (Accessed: 22 July 2022).

Islam, M. A. *et al.* (2019) 'Vitamin D status in patients with systemic lupus erythematosus (SLE): A systematic review and meta-analysis', *Autoimmunity Reviews*, 18(11), p. 102392. doi: 10.1016/j.autrev.2019.102392.

Islam, M. M. *et al.* (2020) 'Use of Mobile Phone App Interventions to Promote Weight Loss: Meta-Analysis', *JMIR mHealth and uHealth*, 8(7), p. e17039. doi: 10.2196/17039.

Israel, C. W. and Manegold, J. C. (2014) 'Elektrischer Sturm', *Herzschrittmachertherapie + Elektrophysiologie*, 25(2), pp. 59–65. doi: 10.1007/s00399-014-0321-y.

Izmailova, E. S., Demanuele, C. and McCarthy, M. (2023) 'Digital health technology derived measures: Biomarkers or clinical outcome assessments?', *Clinical and Translational Science*. doi: 10.1111/cts.13529.

Jakicic, J. M. *et al.* (2016) 'Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss', *JAMA*, 316(11), p. 1161. doi: 10.1001/jama.2016.12858.

Jamshidi, L., Heyvaert, M. and Van den Noortgate, W. (2017) 'Review of Systematic Reviews and Meta-Analyses of Single-Subject Experimental Studies', *AERA Online Paper Repository*, 22(12), pp. 1–11. doi: 10.1111/obr.13331.E-.

Janani, L. *et al.* (2021) 'Effects of Sitagliptin as Monotherapy and Add-On to Metformin on Weight Loss among Overweight and Obese Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis', *Drug Research*, 71(9), pp. 477–488. doi: 10.1055/a-1555-2797.

Jang, J.-P. *et al.* (2020) 'Role of Remote Monitoring in Detection of Atrial Arrhythmia, Stroke Reduction, and Use of Anticoagulation Therapy ― A Systematic Review and Meta-Analysis ―', *Circulation Journal*, 84(11), pp. 1922–1930. doi: 10.1253/circj.CJ-20-0633.

Jeon, E. and Park, H.-A. (2015) 'Nursing Intervention using smartphone technologies; a systematic review and meta-analysis.', *Studies in health technology and informatics*, 210, pp. 321–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25991158.

Jeong, I. C., Bychkov, D. and Searson, P. C. (2019) 'Wearable devices for precision medicine and health state monitoring', *IEEE Transactions on Biomedical Engineering*, 66(5), pp. 1242–1258. doi: 10.1109/TBME.2018.2871638.

Jin, B. *et al.* (2010) 'A meta-analysis of erythropoiesis-stimulating agents in anaemic patients with chronic heart failure', *European Journal of Heart Failure*, 12(3), pp. 249–253. doi: 10.1093/eurjhf/hfp182.

Jin, Y. H. *et al.* (2016) 'A critical appraisal of the methodology and quality of evidence of systematic reviews and meta-analyses of traditional Chinese medical nursing interventions: A systematic review of reviews', *BMJ Open*, 6(11). doi: 10.1136/bmjopen-2016-011514.

Kamei, T. *et al.* (2022) 'The use of wearable devices in chronic disease management to enhance

adherence and improve telehealth outcomes: A systematic review and meta-analysis', *Journal of Telemedicine and Telecare*, 28(5), pp. 342–359. doi: 10.1177/1357633X20937573.

Kang, S.-H. H. *et al.* (2015) 'Cardiac Resynchronization Therapy and QRS Duration: Systematic Review, Meta-analysis, and Meta-regression', *Journal of Korean Medical Science*, 30(1), p. 24. doi: 10.3346/jkms.2015.30.1.24.

Kang, T., Zou, S. and Weng, C. (2019) 'Pretraining to recognize piCO elements from randomized controlled trial literature', *Studies in Health Technology and Informatics*, 264, pp. 188–192. doi: 10.3233/SHTI190209.

Kanitsoraphan, C. *et al.* (2019) 'Baseline fragmented QRS is associated with increased all-cause mortality in heart failure with reduced ejection fraction: A systematic review and meta-analysis', *Annals of Noninvasive Electrocardiology*, 24(2), p. e12597. doi: 10.1111/anec.12597.

Keen, H. I., Pile, K. and Hill, C. L. (2005) 'The prevalence of underpowered randomized clinical trials in rheumatology', *Journal of Rheumatology*, 32(11), pp. 2083–2088.

Khan, M. S. *et al.* (2017) 'Dose of Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers and Outcomes in Heart Failure: A Meta-Analysis', *Circulation: Heart Failure*, 10(8), pp. 1–8. doi: 10.1161/CIRCHEARTFAILURE.117.003956.

Khan, N. *et al.* (2017) 'Digital Health Technologies to Promote Lifestyle Change and Adherence', *Current Treatment Options in Cardiovascular Medicine*, 19(8), p. 60. doi: 10.1007/s11936-017-0560-4.

Kicinski, M., Springate, D. A. and Kontopantelis, E. (2015) 'Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews', *Statistics in Medicine*, 34(20), pp. 2781–2793. doi: 10.1002/sim.6525.

Kirk, M. A. *et al.* (2019) 'Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis', *American Journal of Health Promotion*, 33(5), pp. 778–791. doi: 10.1177/0890117118816278.

Korantzopoulos, P. *et al.* (2009) 'Implantable cardioverter defibrillator therapy in chronic kidney disease: a meta-analysis', *Europace*, 11(11), pp. 1469–1475. doi: 10.1093/europace/eup282.

Kumar, A. and Taggarsi, M. (2021) 'GRADEing quality of evidence and its importance in evidence-based practice', *BMJ Evidence-Based Medicine*, 26(5), pp. 228–230. doi: 10.1136/bmjebm-2020-111525.

Landolina, M. *et al.* (2012) 'Remote Monitoring Reduces Healthcare Use and Improves Quality of Care in Heart Failure Patients With Implantable Defibrillators', *Circulation*, 125(24), pp. 2985–2992. doi: 10.1161/CIRCULATIONAHA.111.088971.

Lau, Y. *et al.* (2020) 'Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials', *Preventive Medicine*, 132(January), p. 106001. doi: 10.1016/j.ypmed.2020.106001.

Lee, Y. H. (2018) 'An overview of meta-analysis for clinicians', *The Korean Journal of Internal Medicine*, 33(2), pp. 277–283. doi: 10.3904/kjim.2016.195.

Lefebvre, C. *et al.* (2019) 'Cochrane Handbook for Systematic Reviews of Interventions, Chapter 4: Searching for and selecting studies. Draft version (29 January 2019).'

Li, L. *et al.* (2021) 'AMSTAR 2 appraisal of systematic reviews and meta-analyses in the field of heart failure from high-impact journals', *In Review*, pp. 1–17. doi: 10.1186/s13643-022-

02029-9.

Li, X. *et al.* (2020) 'Telemonitoring Interventions in COPD Patients: Overview of Systematic Reviews', *BioMed Research International*, 2020. doi: 10.1155/2020/5040521.

Lim, S. *et al.* (2016) 'Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes', *Acta Diabetologica*, 53(2), pp. 189–198. doi: 10.1007/s00592-015-0754-8.

van der Linden, S. D. *et al.* (2021) 'eHealth cognitive rehabilitation for brain tumor patients: results of a randomized controlled trial', *Journal of Neuro-Oncology*, 154(3), pp. 315–326. doi: 10.1007/s11060-021-03828-1.

Lipsmeier, F. *et al.* (2018) 'Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial', *Movement Disorders*, 33(8), pp. 1287–1297. doi: 10.1002/mds.27376.

Liu, Y. *et al.* (2021) 'Association between CRT(D)/ICD and renal insufficiency: A systematic review and meta-analysis', *Seminars in Dialysis*, 34(1), pp. 17–30. doi: 10.1111/sdi.12937.

Long, Y. *et al.* (2021) 'The benefits of defibrillator in heart failure patients with cardiac resynchronization therapy: A meta-analysis', *Pacing and Clinical Electrophysiology*, 44(2), pp. 225–234. doi: 10.1111/pace.14150.

Lorenz, R. C. *et al.* (2019) 'A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool', *Journal of Clinical Epidemiology*, 114, pp. 133–140. doi: 10.1016/j.jclinepi.2019.05.028.

Love, R., Adams, J. and Sluijs, E. M. F. (2019) 'Are school-based physical activity interventions effective and equitable? A meta-analysis of cluster randomized controlled trials with accelerometer-assessed activity', *Obesity Reviews*, 20(6), pp. 859–870. doi: 10.1111/obr.12823.

Lunde, P. *et al.* (2018) 'The Effectiveness of Smartphone Apps for Lifestyle Improvement in Noncommunicable Diseases: Systematic Review and Meta-Analyses', *Journal of Medical Internet Research*, 20(5), p. e162. doi: 10.2196/jmir.9751.

Lüthje, L. *et al.* (2015) 'A randomized study of remote monitoring and fluid monitoring for the management of patients with implanted cardiac arrhythmia devices', *Europace*, 17(8), pp. 1276–1281. doi: 10.1093/europace/euv039.

Lynch, E. A. *et al.* (2019) 'Activity Monitors for Increasing Physical Activity in Adult Stroke Survivors', *Stroke*, 50(1), pp. E4–E5. doi: 10.1161/STROKEAHA.118.023088.

Magno, M. B. *et al.* (2020) 'Associations and risk factors for dental trauma: A systematic review of systematic reviews', *Community Dentistry and Oral Epidemiology*, 48(6), pp. 447–463. doi: 10.1111/cdoe.12574.

Mahajan, R. *et al.* (2020) 'Complex interaction of obesity, intentional weight loss and heart failure: a systematic review and meta-analysis', *Heart*, 106(1), pp. 58–68. doi: 10.1136/heartjnl-2019-314770.

Maharlouei, N. *et al.* (2019) 'The effects of ginger intake on weight loss and metabolic profiles among overweight and obese subjects: A systematic review and meta-analysis of randomized controlled trials', *Critical Reviews in Food Science and Nutrition*, 59(11), pp. 1753–1766. doi: 10.1080/10408398.2018.1427044.

Mahdi Abid, M. *et al.* (2022) 'POSC36 Implantable Cardiac Defibrillators in Heart Failure Patients: An Overview of Reviews and Meta-Analysis', *Value in Health*, 25(1), p. S39. doi:

10.1016/j.jval.2021.11.182.

Mamalaki, E. *et al.* (2022) 'The effectiveness of technology-based interventions for weight loss maintenance: A systematic review of randomized controlled trials with meta-analysis', *Obesity Reviews*, 23(9). doi: 10.1111/obr.13483.

Manuscript, A. (2017) 'Community Activities for Wheelchair Users', 11(5), pp. 361–374. doi: 10.3109/17483107.2014.989420.A.

Martin, D. T. *et al.* (2015) 'Randomized trial of atrial arrhythmia monitoring to guide anticoagulation in patients with implanted defibrillator and cardiac resynchronization devices', *European Heart Journal*, 36(26), pp. 1660–1668. doi: 10.1093/eurheartj/ehv115.

Martin, N. *et al.* (2018) 'Beta-blockers and inhibitors of the renin-angiotensin aldosterone system for chronic heart failure with preserved ejection fraction', *Cochrane Database of Systematic Reviews*, 2021(5). doi: 10.1002/14651858.CD012721.pub3.

Martínez-Mesa, J. *et al.* (2016) 'Sampling: how to select participants in my research study?', *Anais Brasileiros de Dermatologia*, 91(3), pp. 326–330. doi: 10.1590/abd1806-4841.20165254.

Matsui, K. *et al.* (2019) 'Switching to antipsychotic monotherapy vs. staying on antipsychotic polypharmacy in schizophrenia: A systematic review and meta-analysis', *Schizophrenia Research*, 209, pp. 50–57. doi: 10.1016/j.schres.2019.05.030.

McDonough, D. J., Su, X. and Gao, Z. (2021) 'Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis', *British Journal of Sports Medicine*, 55(16), pp. 917–925. doi: 10.1136/bjsports-2020-103594.

McLellan, J. *et al.* (2020) 'Natriuretic peptide-guided treatment for heart failure: A systematic review and meta-analysis', *BMJ Evidence-Based Medicine*, 25(1), pp. 33–37. doi: 10.1136/bmjebm-2019-111208.

MedTech (2015) *Medical devices and pharmaceuticals: Two different worlds in one health setting*. Available at: https://www.medtecheurope.org/news-and-events/news/medical-devices-and-pharmaceuticals-two-different-worlds-in-one-health-setting/ (Accessed: 22 July 2022).

Meerwijk, E. L. *et al.* (2016) 'Direct versus indirect psychosocial and behavioural interventions to prevent suicide and suicide attempts: A systematic review and meta-analysis', *The Lancet Psychiatry*, 3(6), pp. 544–554. doi: 10.1016/S2215-0366(16)00064-X.

Meister, S., Deiters, W. and Becker, S. (2016) 'Digital health and digital biomarkers – enabling value chains on health data', *Current Directions in Biomedical Engineering*, 2(1), pp. 577–581. doi: 10.1515/cdbme-2016-0128.

De Menezes, E. V. A. *et al.* (2019) 'Influence of Paleolithic diet on anthropometric markers in chronic diseases: Systematic review and meta-analysis', *Nutrition Journal*, 18(1). doi: 10.1186/s12937-019-0457-z.

Meng, Y. *et al.* (2017) 'Efficacy of low carbohydrate diet for type 2 diabetes mellitus management: A systematic review and meta-analysis of randomized controlled trials', *Diabetes Research and Clinical Practice*, 131, pp. 124–131. doi: 10.1016/j.diabres.2017.07.006.

Metcalf, B., Henley, W. and Wilkin, T. (2012) 'Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively

measured outcomes (EarlyBird 54)', *BMJ*, 345(sep27 1), pp. e5888–e5888. doi: 10.1136/bmj.e5888.

Mezher, M. (2023) *European Commission moves to delay MDR transition, RF*. Available at: https://www.raps.org/news-and-articles/news-articles/2023/1/european-commission-moves-to-delay-mdr-transition (Accessed: 10 May 2023).

Miller, R. J. H. *et al.* (2015) 'Baseline Functional Class and Therapeutic Efficacy of Common Heart Failure Interventions: A Systematic Review and Meta-analysis', *Canadian Journal of Cardiology*, 31(6), pp. 792–799. doi: 10.1016/j.cjca.2014.12.031.

Min, S. H. *et al.* (2017) 'Comparison between SGLT2 inhibitors and DPP4 inhibitors added to insulin therapy in type 2 diabetes: a systematic review with indirect comparison meta-analysis', *Diabetes/Metabolism Research and Reviews*, 33(1), pp. 1–11. doi: 10.1002/dmrr.2818.

Minaldi, E. *et al.* (2020) 'Thyroid autoimmunity and risk of post-partum depression: a systematic review and meta-analysis of longitudinal studies', *Journal of Endocrinological Investigation*, 43(3), pp. 271–277. doi: 10.1007/s40618-019-01120-8.

Mizuguchi, T. *et al.* (2011) 'Laparoscopic hepatectomy: A systematic review, meta-analysis, and power analysis', *Surgery Today*, 41(1), pp. 39–47. doi: 10.1007/s00595-010-4337-6.

El Moheb, M. *et al.* (2018) 'Implantable cardiac defibrillators for people with non-ischaemic cardiomyopathy', *Cochrane Database of Systematic Reviews*, 2018(12). doi: 10.1002/14651858.CD012738.pub2.

Moll, I. *et al.* (2017) 'Functional electrical stimulation of the ankle dorsiflexors during walking in spastic cerebral palsy: a systematic review', *Developmental Medicine and Child Neurology*, 59(12), pp. 1230–1236. doi: 10.1111/dmcn.13501.

Monaghan, T. F. *et al.* (2021) 'Blinding in Clinical Trials: Seeing the Big Picture', *Medicina*, 57(7), p. 647. doi: 10.3390/medicina57070647.

Moon, S. *et al.* (2021) 'Efficacy and Safety of the New Appetite Suppressant, Liraglutide: A Meta-Analysis of Randomized Controlled Trials', *Endocrinology and Metabolism*, 36(3), pp. 647–660. doi: 10.3803/ENM.2020.934.

Motahari-Nezhad, H. *et al.* (2021) 'Scoping review of systematic reviews of digital biomarker-based studies (Preprint)', *JMIR mHealth and uHealth*. doi: 10.2196/35722.

Motahari-Nezhad, H., Fgaier, M., *et al.* (2022) 'Digital Biomarker-Based Studies: Scoping Review of Systematic Reviews', *JMIR mHealth and uHealth*, 10(10). doi: 10.2196/35722.

Motahari-Nezhad, H., Al-Abdulkarim, H., *et al.* (2022a) 'Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews', *Journal of Medical Internet Research*, 24(12), p. e41042. doi: 10.2196/41042.

Motahari-Nezhad, H., Al-Abdulkarim, H., *et al.* (2022b) 'Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews', *Journal of Medical Internet Research*, 24(12). doi: 10.2196/41042.

Motahari-Nezhad, H. (2023) 'An artificial neural network (ANN) model for publication bias: a machine learning-based study on PubMed meta-analyses', *Aslib Journal of Information Management*. doi: 10.1108/AJIM-08-2022-0364.

Motahari-nezhad, H., Miribonjar, M. and Sadeghdaghighi, A. (2022) 'Methodological and Evidence Synthesis Quality Evaluation of Meta- Analyses Assessing the Effect of Antibacterial Envelopes to Reduce CIED- Related Infections', *International Journal of Medical Research & Health Sciences*, 11(2), pp. 65–76.

Moxey, A. *et al.* (2010) 'Computerized clinical decision support for prescribing: Provision does not guarantee uptake', *Journal of the American Medical Informatics Association*, 17(1), pp. 25–33. doi: 10.1197/jamia.M3170.

Mukherjee, S. *et al.* (2022) 'State of the Art in Smart Portable, Wearable, Ingestible and Implantable Devices for Health Status Monitoring and Disease Management', *Sensors*, 22(11), p. 4228. doi: 10.3390/s22114228.

Mulrow, C. D. (1994) 'Systematic Reviews: Rationale for systematic reviews', *BMJ*, 309(6954), pp. 597–599. doi: 10.1136/bmj.309.6954.597.

Munafò, M. R. *et al.* (2017) 'A manifesto for reproducible science', *Nature Human Behaviour*, 1(1), p. 0021. doi: 10.1038/s41562-016-0021.

Murad, M. H. *et al.* (2018) 'The effect of publication bias magnitude and direction on the certainty in evidence', *BMJ Evidence-Based Medicine*, 23(3), pp. 84–86. doi: 10.1136/bmjebm-2018-110891.

Musgrave, L. *et al.* (2023) 'Addressing Preconception Behavior Change Through Mobile Phone Apps: Systematic Review and Meta-analysis', *Journal of Medical Internet Research*, 25, p. e41900. doi: 10.2196/41900.

Mustafa, U. *et al.* (2018) 'Atrial Fibrillation Is Associated With Higher Overall Mortality in Patients With Implantable Cardioverter-Defibrillator: A Systematic Review and Meta-Analysis', *Journal of the American Heart Association*, 7(22). doi: 10.1161/JAHA.118.010156.

Nair, A. (2019) 'Publication bias - Importance of studies with negative results!', *Indian Journal of Anaesthesia*, 63(6), p. 505. doi: 10.4103/ija.IJA_142_19.

Naka, K. K. *et al.* (2019) 'Association between atrial fibrillation and patient-important outcomes in heart failure patients with implantable cardioverter-defibrillators: a systematic review and meta-analysis', *European Heart Journal - Quality of Care and Clinical Outcomes*, 5(2), pp. 96–104. doi: 10.1093/ehjqcco/qcy054.

Nam, K. H. *et al.* (2019) 'Internet of things, digital biomarker, and artificial intelligence in spine: Current and future perspectives', *Neurospine*, 16(4), pp. 705–711. doi: 10.14245/ns.1938388.194.

Nascimento, S. L. *et al.* (2014) 'The effect of physical exercise strategies on weight loss in postpartum women: a systematic review and meta-analysis', *International Journal of Obesity*, 38(5), pp. 626–635. doi: 10.1038/ijo.2013.183.

Ng, K. T. and Yap, J. L. L. (2018) 'Continuous infusion vs. intermittent bolus injection of furosemide in acute decompensated heart failure: systematic review and meta-analysis of randomised controlled trials', *Anaesthesia*, 73(2), pp. 238–247. doi: 10.1111/anae.14038.

NICE (2019) 'Evidence Standards Framework for Digital Health Technologies'.

NICE (2023) *Evidence standards framework (ESF) for digital health technologies*. Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies (Accessed: 12 March 2023).

Nielsen, E. E. *et al.* (2020) 'Beneficial and harmful effects of sacubitril/valsartan in patients with heart failure: A systematic review of randomised clinical trials with meta-analysis and trial sequential analysis', *Open Heart*, 7(2), pp. 1–15. doi: 10.1136/openhrt-2020-001294.

Nikolaidou, T. *et al.* (2018) 'Postmortem ICD interrogation in mode of death classification', *Journal of Cardiovascular Electrophysiology*, 29(4), pp. 573–583. doi: 10.1111/jce.13414.

Nistor, I. *et al.* (2015) 'Vasopressin receptor antagonists for the treatment of heart failure: a systematic review and meta-analysis of randomized controlled trials', *International Urology and Nephrology*, 47(2), pp. 335–344. doi: 10.1007/s11255-014-0855-2.

Nord, C. L. *et al.* (2017) 'Power-up: A Reanalysis of "Power Failure" in Neuroscience Using Mixture Modeling', *The Journal of Neuroscience*, 37(34), pp. 8051–8061. doi: 10.1523/JNEUROSCI.3592-16.2017.

Nwe, K. *et al.* (2020) 'Medical mobile app classification using the national institute for health and care excellence evidence standards framework for digital health technologies: Interrater reliability study', *Journal of Medical Internet Research*, 22(6). doi: 10.2196/17457.

O'Sullivan, D. *et al.* (2014) 'Decision time for clinical decision support systems', *Clinical Medicine, Journal of the Royal College of Physicians of London*, 14(4), pp. 338–341. doi: 10.7861/clinmedicine.14-4-338.

Otto, L. *et al.* (2020) 'Defining and Delimitating Telemedicine and Related Terms - An Ontology-Based Classification', *Studies in Health Technology and Informatics*, 268, pp. 113–122. doi: 10.3233/SHTI200010.

Oxman, A. D. (1993) 'Users' Guides to the Medical Literature', *JAMA*, 270(17), p. 2093. doi: 10.1001/jama.1993.03510170083036.

Paez, A. (2017) 'Gray literature: An important resource in systematic reviews', *Journal of Evidence-Based Medicine*, 10(3), pp. 233–240. doi: 10.1111/jebm.12266.

Page, M. J. *et al.* (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, p. n71. doi: 10.1136/bmj.n71.

Paltamaa, J. *et al.* (2012) 'Effects of physiotherapy interventions on balance in multiple sclerosis : A systematic review and meta-analysis of randomized controlled trials', *Journal of Rehabilitation Medicine*, 44(10), pp. 811–823. doi: 10.2340/16501977-1047.

Pan, R. *et al.* (2022) 'Effect of SGLT-2 inhibitors on body composition in patients with type 2 diabetes mellitus: A meta-analysis of randomized controlled trials', *PLoS ONE*, 17(12 December), pp. 1–13. doi: 10.1371/journal.pone.0279889.

Paravattil, B., Wilby, K. J. and Turgeon, R. (2016) 'Topiramate monotherapy for weight reduction in patients with type 2 diabetes mellitus: A systematic review and meta-analysis', *Diabetes Research and Clinical Practice*, 114, pp. 9–14. doi: 10.1016/j.diabres.2016.02.002.

Park, S. *et al.* (2019) 'From treatment to prevention: The evolution of digital healthcare', *Nature*, 71(4), pp. 345–376.

Parry, S. *et al.* (2013) 'Participatory Workplace Interventions Can Reduce Sedentary Time for Office Workers—A Randomised Controlled Trial', *PLoS ONE*. Edited by C. P. Earnest, 8(11), p. e78957. doi: 10.1371/journal.pone.0078957.

Paviot, B. T. *et al.* (2011) 'Development of a new international classification of health interventions based on an ontology framework', *Studies in Health Technology and Informatics*, 169, pp. 749–753. doi: 10.3233/978-1-60750-806-9-749.

Payab, M. *et al.* (2020) 'Effect of the herbal medicines in obesity and metabolic syndrome: A systematic review and meta-analysis of clinical trials', *Phytotherapy Research*, 34(3), pp. 526–545. doi: 10.1002/ptr.6547.

Peng, P. *et al.* (2023) 'Effectiveness of Wearable Activity Monitors on Metabolic Outcomes in Patients With Type 2 Diabetes: A Systematic Review and Meta-Analysis', *Endocrine Practice*, 29(5), pp. 368–378. doi: 10.1016/j.eprac.2023.02.004.

Petkovic, J. *et al.* (2021) 'Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population', *Cochrane Database of Systematic Reviews*, 2021(6). doi: 10.1002/14651858.CD012932.pub2.

Pieper, D. *et al.* (2018) 'How is AMSTAR applied by authors – a call for better reporting', *BMC Medical Research Methodology*, 18(1), p. 56. doi: 10.1186/s12874-018-0520-z.

Pigott, T. D. and Polanin, J. R. (2020) 'Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review', *Review of Educational Research*, 90(1), pp. 24–46. doi: 10.3102/0034654319877153.

Pollock, A. *et al.* (2016) 'An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews', *Journal of Clinical Epidemiology*, 70, pp. 106–110. doi: 10.1016/j.jclinepi.2015.08.013.

Pranata, R. *et al.* (2021) 'Clinical frailty scale and mortality in COVID-19: A systematic review and dose-response meta-analysis', *Archives of Gerontology and Geriatrics*, 93(January), p. 104324. doi: 10.1016/j.archger.2020.104324.

Psaltikidis, E. M., Costa, E. A. M. and Graziano, K. U. (2021) 'Reuse of pacemakers and implantable cardioverter-defibrillators: systematic review, meta-analysis and quality assessment of the body of evidence', *Expert Review of Medical Devices*, 18(6), pp. 553–567. doi: 10.1080/17434440.2021.1927706.

Qian, C. *et al.* (2016) 'The Efficacy and Safety of Iron Supplementation in Patients With Heart Failure and Iron Deficiency: A Systematic Review and Meta-analysis', *Canadian Journal of Cardiology*, 32(2), pp. 151–159. doi: 10.1016/j.cjca.2015.06.009.

Rabar, S. *et al.* (2012) 'Risk assessment of fragility fractures: summary of NICE guidance', *BMJ*, 345(aug08 1), pp. e3698–e3698. doi: 10.1136/bmj.e3698.

Rattanawong, P. *et al.* (2018) 'Atrial fibrillation is associated with sudden cardiac death: a systematic review and meta-analysis', *Journal of Interventional Cardiac Electrophysiology*, 51(2), pp. 91–104. doi: 10.1007/s10840-017-0308-9.

Rehm, J. *et al.* (2017) 'The relationship between different dimensions of alcohol use and the burden of disease—an update', *Addiction*, 112(6), pp. 968–1001. doi: 10.1111/add.13757.

Richardson, C. R. *et al.* (2008) 'A Meta-Analysis of Pedometer-Based Walking Interventions and Weight Loss', *The Annals of Family Medicine*, 6(1), pp. 69–77. doi: 10.1370/afm.761.

Ringeval, M. *et al.* (2020) 'Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis', *Journal of Medical Internet Research*, 22(10), p. e23954. doi: 10.2196/23954.

Robb, M. A., McInnes, P. M. and Califf, R. M. (2016) 'Biomarkers and Surrogate Endpoints', *JAMA*, 315(11), p. 1107. doi: 10.1001/jama.2016.2240.

Rogers, J. P. *et al.* (2020) 'Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic', *The Lancet Psychiatry*, 7(7), pp. 611–627. doi: 10.1016/S2215-0366(20)30203-0.

Rorris, F.-P. *et al.* (2021) 'Implantable cardioverter defibrillators in left ventricular assist device patients: A systematic review and meta-analysis', *The Journal of Heart and Lung Transplantation*, 40(10), pp. 1098–1106. doi: 10.1016/j.healun.2021.05.014.

Rossi, A. *et al.* (2008) 'The current role of cardiac resynchronization therapy in reducing mortality and hospitalization in heart failure patients: a meta-analysis from clinical trials',

Heart and Vessels, 23(4), pp. 217–223. doi: 10.1007/s00380-008-1039-8.

Rote, A. E. (2017) 'Physical activity intervention using Fitbits in an introductory college health course', *Health Education Journal*, 76(3), pp. 337–348. doi: 10.1177/0017896916674505.

Ryu, B. *et al.* (2017) 'Impact of an Electronic Health Record-Integrated Personal Health Record on Patient Participation in Health Care: Development and Randomized Controlled Trial of MyHealthKeeper', *Journal of Medical Internet Research*, 19(12), p. e401. doi: 10.2196/jmir.8867.

Saeteaw, M. *et al.* (2021) 'Efficacy and safety of pharmacological cachexia interventions: systematic review and network meta-analysis', *BMJ Supportive & Palliative Care*, 11(1), pp. 75–85. doi: 10.1136/bmjspcare-2020-002601.

Sanghera, J. *et al.* (2020) 'The impact of SARS-CoV-2 on the mental health of healthcare workers in a hospital setting—A Systematic Review', *Journal of Occupational Health*, 62(1), pp. 1–16. doi: 10.1002/1348-9585.12175.

Schardt, C. *et al.* (2007) 'Utilization of the PICO framework to improve searching PubMed for clinical questions', *BMC Medical Informatics and Decision Making*, 7, pp. 1–6. doi: 10.1186/1472-6947-7-16.

Schock, D., Neher, J. O. and Safranek, S. (2017) 'Clinical Inquiry: Do pedometers increase activity and improve health outcomes?', *The Journal of family practice*, 66(1), pp. 48–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28188319.

Schuch, F. B. *et al.* (2016) 'Exercise for depression in older adults: A meta-analysis of randomized controlled trials adjusting for publication bias', *Revista Brasileira de Psiquiatria*, 38(3), pp. 247–254. doi: 10.1590/1516-4446-2016-1915.

Schünemann, H. *et al.* (2013) *Introduction to GRADE Handbook Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach*. Available at: https://gdt.gradepro.org/app/handbook/handbook.html#h.9rdbelsnu4iy (Accessed: 15 January 2021).

Sedgwick, P. (2015) 'Meta-analyses: what is heterogeneity?', *BMJ*, 350(mar16 1), pp. h1435–h1435. doi: 10.1136/bmj.h1435.

Serdar, C. C. *et al.* (2021) 'Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies', *Biochemia medica*, 31(1), pp. 27–53. doi: 10.11613/BM.2021.010502.

Sfairopoulos, D. *et al.* (2022) 'Association between sodium-glucose cotransporter-2 inhibitors and risk of sudden cardiac death or ventricular arrhythmias: A meta-analysis of randomized controlled trials', *Europace*, 24(1), pp. 20–30. doi: 10.1093/europace/euab177.

Shah, R. M. *et al.* (2014) 'Cardiac-resynchronization therapy in patients with systolic heart failure and QRS interval ≤130 ms: Insights from a meta-analysis', *Europace*, 17(2), pp. 267–273. doi: 10.1093/europace/euu214.

Shaw, J. *et al.* (2018) 'Beyond "implementation": digital health innovation and service design', *npj Digital Medicine*, 1(1), pp. 1–5. doi: 10.1038/s41746-018-0059-8.

Shea, B. J. *et al.* (2009) 'AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews', *Journal of Clinical Epidemiology*, 62(10), pp. 1013–1020. doi: 10.1016/j.jclinepi.2008.10.009.

Shea, B. J. *et al.* (2017a) 'AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both', *BMJ*, 358,

p. j4008. doi: 10.1136/bmj.j4008.

Shea, B. J. *et al.* (2017b) 'AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both', *BMJ (Online)*, 358, pp. 1–9. doi: 10.1136/bmj.j4008.

Shen, L. F. *et al.* (2018) 'Effect of roflumilast on chronic obstructive pulmonary disease: a systematic review and meta-analysis', *Irish Journal of Medical Science*, 187(3), pp. 731–738. doi: 10.1007/s11845-018-1738-9.

Shin, E. K. *et al.* (2018) 'Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits', *npj Digital Medicine*, 1(1), p. 50. doi: 10.1038/s41746-018-0056-y.

Shun-Shin, M. J. *et al.* (2017) 'Implantable cardioverter defibrillators for primary prevention of death in left ventricular dysfunction with and without ischaemic heart disease: A meta-analysis of 8567 patients in the 11 trials', *European Heart Journal*, 38(22), pp. 1738–1746. doi: 10.1093/eurheartj/ehx028.

Shurrab, M. *et al.* (2018) 'Outcomes of ICDs and CRTs in patients with chronic kidney disease: a meta-analysis of 21,000 patients', *Journal of Interventional Cardiac Electrophysiology*, 53(1), pp. 123–129. doi: 10.1007/s10840-018-0424-1.

Siddiqui, W. J. *et al.* (2018) 'Prophylactic use of the implantable cardioverter-defibrillator and its effect on the long-term survival, cardiovascular and sudden cardiac death in nonischemic cardiomyopathy patients—a systematic review and meta-analysis', *Heart Failure Reviews*, 23(2), pp. 181–190. doi: 10.1007/s10741-018-9671-6.

Silva, A. M. *et al.* (2018) 'What is the effect of diet and/or exercise interventions on behavioural compensation in non-exercise physical activity and related energy expenditure of free-living adults? A systematic review', *British Journal of Nutrition*, 119(12), pp. 1327–1345. doi: 10.1017/S000711451800096X.

Sipahi, I. *et al.* (2015) 'Impact of QRS duration on survival benefit with prophylactic implantable cardioverter-defibrillators', *Journal of Cardiovascular Medicine*, 16(12), pp. 811–816. doi: 10.2459/JCM.0b013e3283657524.

Sjögren, P. and Hedström, L. (2010) 'Sample size determination and statistical power in randomized controlled trials', *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, 109(5), pp. 652–653. doi: 10.1016/j.tripleo.2010.01.013.

Smith, S. M. *et al.* (2021) 'Telehealth in surgery: an umbrella review', *ANZ Journal of Surgery*, 91(11), pp. 2360–2375. doi: 10.1111/ans.17217.

Snowball, R., Library, C. and Hospital, J. R. (1997) 'Snowball-1997-Health_Libraries_Review', pp. 167–172.

Soares, R. C. *et al.* (2021) 'Methods for prevention of early childhood caries: Overview of systematic reviews', *International Journal of Paediatric Dentistry*, 31(3), pp. 394–421. doi: 10.1111/ipd.12766.

Song, G. M. *et al.* (2016) 'Regime for bowel preparation in patients scheduled to colonoscopy: Low-residue diet or clear liquid diet? evidence from systematic review with power analysis', *Medicine (United States)*, 95(1), pp. 1–12. doi: 10.1097/MD.0000000000002432.

Song, X. *et al.* (2017) 'Efficacy and Safety of L-Carnitine Treatment for Chronic Heart Failure: A Meta-Analysis of Randomized Controlled Trials', *BioMed Research International*, 2017. doi: 10.1155/2017/6274854.

Sood, S. *et al.* (2007) 'What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings', *Telemedicine and e-Health*, 13(5), pp. 573–590. doi: 10.1089/tmj.2006.0073.

Sovacool, B. K. and Hess, D. J. (2017) 'Ordering theories: Typologies and conceptual frameworks for sociotechnical change', *Social Studies of Science*, 47(5), pp. 703–750. doi: 10.1177/0306312717709363.

STAVRAKIS, S., PATEL, N. H. and REYNOLDS, D. W. (2013) 'Defibrillation Threshold Testing Does Not Predict Clinical Outcomes during Long-Term Follow-Up: A Meta-Analysis', *Pacing and Clinical Electrophysiology*, 36(11), pp. 1402–1408. doi: 10.1111/pace.12218.

Steel, N. *et al.* (2014) 'A review of clinical practice guidelines found that they were often based on evidence of uncertain relevance to primary care patients', *Journal of Clinical Epidemiology*, 67(11), pp. 1251–1257. doi: 10.1016/j.jclinepi.2014.05.020.

Stevens, S. K. *et al.* (1996) 'Effects of oral propafenone on defibrillation and pacing thresholds in patients receiving implantable cardioverter-defibrillators', *Journal of the American College of Cardiology*, 28(2), pp. 418–422. doi: 10.1016/0735-1097(96)00156-8.

Stevenson, J. K. *et al.* (2019) 'Ehealth interventions for people with chronic kidney disease', *Cochrane Database of Systematic Reviews*, 2019(8). doi: 10.1002/14651858.CD012379.pub2.

Sullivan, G. M. and Feinn, R. (2012) 'Using Effect Size—or Why the P Value Is Not Enough', *Journal of Graduate Medical Education*, 4(3), pp. 279–282. doi: 10.4300/JGME-D-12-00156.1.

Sun, W. P. *et al.* (2016) 'Long-term efficacy of implantable cardiac resynchronization therapy plus defibrillator for primary prevention of sudden cardiac death in patients with mild heart failure: an updated meta-analysis', *Heart Failure Reviews*, 21(4), pp. 447–453. doi: 10.1007/s10741-016-9550-y.

Sutton, A. J. and Higgins, J. P. T. (2008) 'Recent developments in meta-analysis', *Statistics in Medicine*, 27(5), pp. 625–650. doi: 10.1002/sim.2934.

Szajewska, H. (2018a) 'Evidence-Based Medicine and Clinical Research: Both Are Needed, Neither Is Perfect', *Annals of Nutrition and Metabolism*, 72(suppl 3), pp. 13–23. doi: 10.1159/000487375.

Szajewska, H. (2018b) 'Evidence-Based Medicine and Clinical Research: Both Are Needed, Neither Is Perfect', *Annals of Nutrition and Metabolism*, 72(Suppl. 3), pp. 13–23. doi: 10.1159/000487375.

Sze, E. and Daubert, J. P. (2015) 'Why the Authors Use Cardiac Resynchronization Therapy with Defibrillators', *Cardiac Electrophysiology Clinics*, 7(4), pp. 695–707. doi: 10.1016/j.ccep.2015.08.017.

Talenezhad, N. *et al.* (2020) 'Effects of L-carnitine supplementation on weight loss and body composition: A systematic review and meta-analysis of 37 randomized controlled clinical trials with dose-response analysis', *Clinical Nutrition ESPEN*, 37(xxxx), pp. 9–23. doi: 10.1016/j.clnesp.2020.03.008.

Tang, M. S. S. *et al.* (2020) 'Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials', *JMIR mHealth and uHealth*, 8(7), p. e15576. doi: 10.2196/15576.

Tebala, G. D. (2015) 'What is the future of biomedical research?', *Medical Hypotheses*, 85(4), pp. 488–490. doi: 10.1016/j.mehy.2015.07.003.

Terrin, N., Schmid, C. H. and Lau, J. (2005) 'In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias', *Journal of Clinical Epidemiology*, 58(9), pp. 894–901. doi: 10.1016/j.jclinepi.2005.01.006.

The Lancet (2018) 'Is digital medicine different?', *The Lancet*, 392(10142), p. 95. doi: 10.1016/S0140-6736(18)31562-9.

The Lancet (2019) 'Icd-11', *The Lancet*, 393(10188), p. 2275. doi: 10.1016/S0140-6736(19)31205-X.

Thorlund, K. *et al.* (2011) 'The Number of Patients and Events Required to Limit the Risk of Overestimation of Intervention Effects in Meta-Analysis—A Simulation Study', *PLoS ONE*. Edited by G. Biondi-Zoccai, 6(10), p. e25491. doi: 10.1371/journal.pone.0025491.

Thorndike, A. N. *et al.* (2014) 'Activity Monitor Intervention to Promote Physical Activity of Physicians-In-Training: Randomized Controlled Trial', *PLoS ONE*. Edited by N. Hayashi, 9(6), p. e100251. doi: 10.1371/journal.pone.0100251.

Toledo, P. R. *et al.* (2023) 'Interpersonal psychotherapy for treatment of obesity: A systematic review and meta-analysis', *Journal of Affective Disorders*, 320(September 2022), pp. 319–329. doi: 10.1016/j.jad.2022.09.070.

Tseng, A. S. *et al.* (2019) 'Efficacy of Pharmacologic and Cardiac Implantable Electronic Device Therapies in Patients With Heart Failure and Reduced Ejection Fraction', *Circulation: Arrhythmia and Electrophysiology*, 12(6). doi: 10.1161/CIRCEP.118.006951.

Tu, R. *et al.* (2013) '[A systematic review and meta-analysis on efficacy and safety of cardiac resynchronization therapy alone or in combination with implantable cardioversion defibrillation in patients with mild to severe heart failure].', *Zhonghua xin xue guan bing za zhi*, 41(2), pp. 161–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23710749.

Unsworth, H. *et al.* (2021) 'The NICE Evidence Standards Framework for digital health and care technologies – Developing and maintaining an innovative evidence framework with global impact', *Digital Health*, 7, pp. 1–20. doi: 10.1177/20552076211018617.

Vakil, K. *et al.* (2016) 'Implantable Cardioverter-Defibrillator Use in Patients With Left Ventricular Assist Devices', *JACC: Heart Failure*, 4(10), pp. 772–779. doi: 10.1016/j.jchf.2016.05.003.

Vasudevan, S. *et al.* (2022) 'Digital biomarkers: Convergence of digital health technologies and biomarkers', *npj Digital Medicine*, 5(1), pp. 3–5. doi: 10.1038/s41746-022-00583-z.

De Vecchis, R. *et al.* (2015) 'Besseres renales Sicherheitsprofil und klinische Ergebnisse durch hypertone Kochsalzlösung plus Furosemid i.v. bei akuter dekompensierter Herzinsuffizienz: Eine Metaanalyse der Literatur', *Herz*, 40(3), pp. 423–435. doi: 10.1007/s00059-013-4041-6.

Veronese, N. *et al.* (2019) 'Is chocolate consumption associated with health outcomes? An umbrella review of systematic reviews and meta-analyses', *Clinical Nutrition*, 38(3), pp. 1101–1108. doi: 10.1016/j.clnu.2018.05.019.

Vettoretti, M. *et al.* (2020) 'Advanced Diabetes Management Using Artificial Intelligence and Continuous Glucose Monitoring Sensors', *Sensors*, 20(14), p. 3870. doi: 10.3390/s20143870.

Villarreal-Zegarra, D. *et al.* (2022) 'Development of a Framework for the Implementation of Synchronous Digital Mental Health: Realist Synthesis of Systematic Reviews', *JMIR Mental Health*, 9(3), p. e34760. doi: 10.2196/34760.

Villinger, K. *et al.* (2019) 'The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis',

Obesity Reviews, 20(10), pp. 1465–1484. doi: 10.1111/obr.12903.

Vitger, T. *et al.* (2021) 'Digital Shared Decision-Making Interventions in Mental Healthcare: A Systematic Review and Meta-Analysis', *Frontiers in Psychiatry*, 12(September). doi: 10.3389/fpsyt.2021.691251.

Vyas, S. and Bhargava, D. (2021) *Smart Health Systems*, *Smart Health Systems*. Singapore: Springer Singapore. doi: 10.1007/978-981-16-4201-2.

Wang, C., Xiong, B. and Cai, L. (2017) 'Effects of Tolvaptan in patients with acute heart failure: A systematic review and meta-analysis', *BMC Cardiovascular Disorders*, 17(1), pp. 1–11. doi: 10.1186/s12872-017-0598-y.

Wang, M., Zheng, Y. and Jin, H. (2021) 'Ultrafiltration for patients with acute decompensated heart failure', *Medicine*, 100(50), p. e28029. doi: 10.1097/md.0000000000028029.

Wang, R. *et al.* (2022) 'Effect of Sacubitril/Valsartan on Reducing the Risk of Arrhythmia: A Systematic Review and Meta-Analysis of Randomized Controlled Trials', *Frontiers in Cardiovascular Medicine*, 9(July). doi: 10.3389/fcvm.2022.890481.

Wang, S.-A. *et al.* (2020) 'Effects of real-time feedback on cardiopulmonary resuscitation quality on outcomes in adult patients with cardiac arrest: A systematic review and meta-analysis', *Resuscitation*, 155(252), pp. 82–90. doi: 10.1016/j.resuscitation.2020.07.024.

Wang, W. *et al.* (2022) 'The Effectiveness of Wearable Devices as Physical Activity Interventions for Preventing and Treating Obesity in Children and Adolescents: Systematic Review and Meta-analysis', *JMIR mHealth and uHealth*, 10(4), pp. 1–13. doi: 10.2196/32435.

Wasiak, M. *et al.* (2023) 'An implantable cardioverter-defibrillator for primary prevention in non-ischemic cardiomyopathy: A systematic review and meta-analysis', *Cardiology Journal*, 30(1), pp. 117–124. doi: 10.5603/CJ.a2021.0041.

Whiting, P. *et al.* (2003) 'The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews', *BMC Medical Research Methodology*, 3(1), p. 25. doi: 10.1186/1471-2288-3-25.

Whiting, P. *et al.* (2016) 'ROBIS: A new tool to assess risk of bias in systematic reviews was developed', *Journal of Clinical Epidemiology*, 69, pp. 225–234. doi: 10.1016/j.jclinepi.2015.06.005.

WHO (2017a) *International Classification of Functioning, Disability and Health (ICF)*. Available at: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health (Accessed: 19 August 2022).

WHO (2017b) *International Classification of Functioning, Disability and Health (ICF)*.

WHO (2020a) *ICD-11 International Classification of Diseases 11th Revision The global standard for diagnostic health information*. Available at: https://icd.who.int/en (Accessed: 12 September 2020).

WHO (2020b) *No International Classification of Health Interventions (ICHI)*. Available at: https://www.who.int/standards/classifications/international-classification-of-health-interventions (Accessed: 17 August 2022).

WHO Guideline (1980) *Food and Nutrition Bulletin*. doi: 10.1177/156482658000200103.

Wiersma, R. *et al.* (2020) 'Unravelling the association between accelerometer-derived physical activity and adiposity among preschool children: A systematic review and meta-analyses', *Obesity Reviews*, 21(2). doi: 10.1111/obr.12936.

Wolff, G. *et al.* (2017) 'Implantable cardioverter/defibrillators for primary prevention in dilated cardiomyopathy post-DANISH: an updated meta-analysis and systematic review of randomized controlled trials', *Clinical Research in Cardiology*, 106(7), pp. 501–513. doi: 10.1007/s00392-017-1079-0.

Wong, S. H. *et al.* (2022) 'Wearable technology-delivered lifestyle intervention amongst adults with overweight and obese: A systematic review and meta-regression', *International Journal of Nursing Studies*, 127. doi: 10.1016/j.ijnurstu.2021.104163.

Wong, W. C., Cheung, C. S. and Hart, G. J. (2008) 'Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours', *Emerging Themes in Epidemiology*, 5(1), p. 23. doi: 10.1186/1742-7622-5-23.

Woods, B. *et al.* (2015) 'Individual patient data network meta-analysis of mortality effects of implantable cardiac devices', *Heart*, 101(22), pp. 1800–1806. doi: 10.1136/heartjnl-2015-307634.

Woods, S. P. *et al.* (2006) 'Statistical power of studies examining the cognitive effects of subthalamic nucleus deep brain stimulation in Parkinson's disease', *Clinical Neuropsychologist*, 20(1), pp. 27–38. doi: 10.1080/13854040500203290.

World Health Organization (2020) *International Classification of Health Interventions (ICHI)*. Available at: https://www.who.int/standards/classifications/international-classification-of-health-interventions (Accessed: 10 September 2020).

Wu, J. *et al.* (2022) 'Metabolic effects of the dual SGLT 1/2 inhibitor sotagliflozin on blood pressure and body weight reduction in people with diabetes: An updated meta-analysis of randomized controlled trials', *Journal of Diabetes and its Complications*, 36(12), p. 108352. doi: 10.1016/j.jdiacomp.2022.108352.

Xing, Z. *et al.* (2017) 'Effectiveness of implantation of cardioverter-defibrillators therapy in patients with non-ischemic heart failure: An updated systematic review and meta-analysis', *Brazilian Journal of Cardiovascular Surgery*, 32(5), pp. 417–422. doi: 10.21470/1678-9741-2017-0003.

Xiong, B. *et al.* (2015) 'The short-term and long-term effects of tolvaptan in patients with heart failure: a meta-analysis of randomized controlled trials', *Heart Failure Reviews*, 20(6), pp. 633–642. doi: 10.1007/s10741-015-9503-x.

Yang, Q. *et al.* (2022) 'The effects of low-fat, high-carbohydrate diets vs. low-carbohydrate, high-fat diets on weight, blood pressure, serum liquids and blood glucose: a systematic review and meta-analysis', *European Journal of Clinical Nutrition*, 76(1), pp. 16–27. doi: 10.1038/s41430-021-00927-0.

Yang, Y. *et al.* (2023) 'Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology', *BMC Biology*, 21(1), p. 71. doi: 10.1186/s12915-022-01485-y.

Yassir, Y. A., McIntyre, G. T. and Bearn, D. R. (2021) 'Orthodontic treatment and root resorption: An overview of systematic reviews', *European Journal of Orthodontics*, 43(4), pp. 442–456. doi: 10.1093/ejo/cjaa058.

Yen, Hsin-Yen Yen and Chiu, H. L. (2019) 'The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials', *Obesity Reviews*, 20(10), pp. 1485–1493. doi: 10.1111/obr.12909.

Yen, Hsin Yen and Chiu, H. L. (2019) 'The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials', *Obesity Reviews*, 20(10), pp. 1485–1493. doi: 10.1111/obr.12909.

Yen, H. Y., Jin, G. and Chiu, H. L. (2023) 'Smartphone app-based interventions targeting physical activity for weight management: A meta-analysis of randomized controlled trials', *International Journal of Nursing Studies*, 137, p. 104384. doi: 10.1016/j.ijnurstu.2022.104384.

Yuan, X. *et al.* (2020) 'Effect of the ketogenic diet on glycemic control, insulin resistance, and lipid metabolism in patients with T2DM: a systematic review and meta-analysis', *Nutrition and Diabetes*, 10(1). doi: 10.1038/s41387-020-00142-z.

Yuan, Y. and Hunt, R. H. (2009) 'Systematic Reviews: The Good, the Bad and the Ugly', *The American Journal of Gastroenterology*, 104(5), pp. 1086–1092. doi: 10.1038/ajg.2009.118.

Yue, T. *et al.* (2020) 'Prognostic Value of Late Gadolinium Enhancement in Predicting Life-Threatening Arrhythmias in Heart Failure Patients With Implantable Cardioverter-Defibrillators: A Systematic Review and Meta-Analysis', *Journal of Magnetic Resonance Imaging*, 51(5), pp. 1422–1439. doi: 10.1002/jmri.26982.

Yuyun, M. F. *et al.* (2021) 'Risk of ventricular arrhythmia in cardiac resynchronization therapy responders and super-responders: a systematic review and meta-analysis', *EP Europace*, 23(8), pp. 1262–1274. doi: 10.1093/europace/euaa414.

Zah, V. *et al.* (2022) 'Paying for Digital Health Interventions – What Evidence is Needed?', *Acta Polytechnica Hungarica*, 19(9), pp. 179–199. doi: 10.12700/APH.19.9.2022.9.10.

Zhang, Q., Zhou, Y. and Yu, C.-M. (2015) 'Incidence, definition, diagnosis, and management of the cardiac resynchronization therapy nonresponder', *Current Opinion in Cardiology*, 30(1), pp. 40–49. doi: 10.1097/HCO.0000000000000140.

Zhang, Yuan *et al.* (2019) 'GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences—inconsistency, imprecision, and other domains', *Journal of Clinical Epidemiology*, 111, pp. 83–93. doi: 10.1016/j.jclinepi.2018.05.011.

Zheng, S. L. *et al.* (2018) 'Drug treatment effects on outcomes in heart failure with preserved ejection fraction: A systematic review and meta-analysis', *Heart*, 104(5), pp. 407–415. doi: 10.1136/heartjnl-2017-311652.

Zhong, P. *et al.* (2022) 'Efficacy and safety of once-weekly semaglutide in adults with overweight or obesity: a meta-analysis', *Endocrine*, 75(3), pp. 718–724. doi: 10.1007/s12020-021-02945-1.

Zhou, C. *et al.* (2022) 'Ketogenic Diet Benefits to Weight Loss, Glycemic Control, and Lipid Profiles in Overweight Patients with Type 2 Diabetes Mellitus: A Meta-Analysis of Randomized Controlled Trails', *International Journal of Environmental Research and Public Health*, 19(16). doi: 10.3390/ijerph191610429.

Zhou, L. *et al.* (2022) 'Effects of e-health interventions on health outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis', *Journal of Clinical Nursing*, (December 2021), pp. 1–12. doi: 10.1111/jocn.16236.

Zhou, Z. *et al.* (2021) 'Effects of music-based movement therapy on motor function, balance, gait, mental health, and quality of life for patients with Parkinson's disease: A systematic review and meta-analysis', *Clinical Rehabilitation*, 35(7), pp. 937–951. doi: 10.1177/0269215521990526.

Zrubka, Z. *et al.* (2021) ‘PP424 Piloting A Comprehensive Search For eHealth Definitions In The Grey Literature: Preliminary Results From A Systematic Scoping Review’, *International Journal of Technology Assessment in Health Care*, 37(S1), pp. 36–36. doi: 10.1017/S0266462321001562.

Appendix I. Statistical power calculations

Statistical power of single studies

Formula 1	Pooled SD= $\sqrt{}$ $((n_1-1)SD_1^2+(n_2-1)SD_2^2)/(n_1+n_2-2)$ (Higgins *et al.*, 2019b)
Formula 2	Cohen's d = m_1-m_2/Pooled SD
Formula 3	pwr.t2n.test(n_1 =, n_2=, d =, sig.level = 0.05)
Formula 4	pwr.t.test(n = , d =, sig.level = 0.05)
Formula 5	power twoproportions p_1 p_2, n() alpha (0.05)

Formula 1: n_1 and n_2 are the sample sizes in treatment and control groups. SD_1 and SD_2 denote the standard deviation in treatment and control group.

Formula 2: m_1 and m_2 represent the mean in treatment and control groups.

Formula 3: n_1 and n_2 are the sample sizes in treatment and control groups. d represents the calculated Cohen's d using the formula 2.

Formula 4: n denotes the total sample size.

Formula 5: p_1 and p_2 represent the proportion of events in treatment and control group and n denotes the total samples.

Statistical power of meta-analyses

Formula number	Type of effect size	Formula used
Formula 6	Odds ratio	power.analysis(OR=,k=,n_1=,n_1=,p=0.05,heterogeneity=)
Formula 7	Risk ratio	First the risk ratio converted to Odds ratio and then formula 6 was used to calculate the power.

Formula 8	Hazard ratio	es <- # summary effect size
		as <- # Average per number per group
		mk <- # Number of included studies
		hg <- # Heterogeniety (".33" for small, "1" for moderate, & "3" for large)
		eq1 <- ((as+as)/((as)*(as))) + ((es^2)/(2*(as+as)))
		eq2 <- hg*(eq1)
		eq3 <- eq2+eq1
		eq4 <- eq3/mk
		eq5 <- (es/sqrt(eq4))
		Power <- (1-pnorm(1.96-eq5)) # Two-tailed
		Power

Formula 9	Continuous	power.analysis(d=,k=,n_1=,n_2=,p=0.05,heterogeneity=)

Formula 6: OR: odds ratio, K: number of included studies, n_1: number of samples in treatment group, n_2: number of samples in control group, heterogeneity: “small”, or “moderate”, or “high” (Higgins, 2003).

Formula 9: d: Cohen’s d effect size, k: number of included studies, n_1: number of samples in treatment group, n_2: number of samples in control group, heterogeneity: “small”, or “moderate”, or “high” (Higgins, 2003).

Appendix II. Search strategies, the characteristics of the included and excluded systematic reviews and their matched non-DBMs or drug-based peers

Search strategy for finding implantable cardiac defibrillator-based meta-analyses in heart failure patients for mortality

Search strategy	Number of identified studies
(("implantable cardiac defibrillator"[Title/Abstract] OR "defibrillators, implantable"[MeSH Terms]) AND ("heart failure"[Title/Abstract] OR "heart failure"[MeSH Terms]) AND ("all-cause mortality"[Title/Abstract] OR "mortality"[MeSH Terms]) AND "meta analysis"[Publication Type])	67

The list of included ICD-based systematic reviews.

Study	Title
(Alba, Foroutan, *et al.*, 2018)	Implantable cardiac defibrillator and mortality in nonischaemic cardiomyopathy: an updated meta-analysis
(Huang *et al.*, 2010)	All cause mortality of cardiac resynchronization therapy with implantable cardioverter defibrillator: A meta-analysis of randomized controlled trials
(Bertoldi *et al.*, 2011)	Mortality Reduction of Cardiac Resynchronization and Implantable Cardioverter-Defibrillator Therapy in Heart Failure: An Updated Meta-Analysis. Does Recent Evidence Change the Standard of Care?
(Kang *et al.*, 2015)	Cardiac Resynchronization Therapy and QRS Duration: Systematic Review, Meta-analysis, and Meta-regression
(Shah *et al.*, 2014)	Cardiac-resynchronization therapyinpatientswith systolic heart failure andQRS interval ≤130ms: insights from ameta-analysis
(Shaojie Chen *et al.*, 2013)	The efficacy and safety of cardiac resynchronization therapy combined with implantable cardioverter defibrillator for heart failure: a meta-analysis of 5674 patients
(Xing *et al.*, 2017)	Effectiveness of Implantation of Cardioverter- Defibrillators Therapy in Patients with Non-Ischemic Heart Failure: an Updated Systematic Review and Meta-Analysis
(Sun *et al.*, 2016)	Long-term efficacy of implantable cardiac resynchronization therapy plus defibrillator for primary prevention of sudden cardiac death in patients with mild heart failure: an updated meta-analysis
(Shun-Shin *et al.*, 2017)	Implantable cardioverter defibrillators for primary prevention of death in left ventricular dysfunction with and without ischaemic heart disease: ameta-analysis of 8567 patients in the 11 trials
(Gama *et al.*, 2020)	Implantable Cardioverter–Defibrillators in Trials of Drug Therapy for Heart Failure: A Systematic Review and Meta-Analysis
(Akel and Lafferty, 2017)	Implantable cardioverter defibrillators for primary prevention in patients with nonischemic cardiomyopathy: A systematic review and meta-analysis

(Barakat *et al.*, 2017)	Primary prevention implantable cardioverter defibrillator in patients with non-ischaemic cardiomyopathy: a meta-analysis of randomized controlled trials
(Alotaibi *et al.*, 2020)	Remote monitoring of implantable cardiac devices in heart failure patients: a systematic review and meta-analysis of randomized controlled trials
(Al-Khatib *et al.*, 2017)	Primary Prevention Implantable Cardioverter Defibillators in Patients With Nonischemic Cardiomyopathy A Meta analysis
(Miller *et al.*, 2015)	Baseline Functional Class and Therapeutic Efficacy of Common Heart Failure Interventions: A Systematic Review and Meta-analysis
(Siddiqui *et al.*, 2018)	Prophylactic use of the implantable cardioverter-defibrillator and its effect on the long-term survival, cardiovascular and sudden cardiac death in nonischemic cardiomyopathy patients—a systematic review and meta-analysis
(Wolff *et al.*, 2017)	Implantable cardioverter/defibrillators for primary prevention in dilated cardiomyopathy post-DANISH: an updated meta analysis and systematic review of randomized controlled trials
(El Moheb *et al.*, 2018)	Implantable cardiac defibrillators for people with non-ischaemic cardiomyopathy (Review)
(Ghanbari, 2009)	Effectiveness of Implantable Cardioverter- Defibrillators for the Primary Prevention of Sudden Cardiac Death in Women With Advanced Heart Failure
(Abdulla, Haarbo, *et al.*, 2006)	Impact of Implantable Defi brillators and Resynchronization Therapy on Outcome in Patients with Left Ventricular Dysfunction – A Meta-Analysis

The list of excluded ICD-based systematic reviews

Study	Title
(Liu *et al.*, 2021)	Association between CRT(D)/ICD and renal insufficiency: A systematic review and meta-analysis
(AlTurki *et al.*, 2019)	Implantable cardioverter-defibrillator use in elderly patients receiving cardiac resynchronization: A meta-analysi
(Bazoukis *et al.*, 2019)	Impact of Implantable Cardioverter-Defibrillator Interventions on All-Cause Mortality in Heart Failure Patients: A Meta-Analysis
(Woods *et al.*, 2015)	Individual patient data network meta-analysis of mortality effects of implantable cardiac devices
(Rattanawong *et al.*, 2018)	Atrial fibrillation is associated with sudden cardiac death: a systematic review and meta-analysis
(Daubert, Martins and Leclercq, 2015)	Why We Have to Use Cardiac Resynchronization Therapy-Pacemaker More
(Engstrom *et al.*, 2022)	Fragmented QRS is associated with ventricular arrhythmias in heart failure patients: A systematic review and meta-analysis
(Alba, Walter, *et al.*, 2018)	Predicting Survival in Patients With Heart Failure With an Implantable Cardioverter Defibrillator: The Heart Failure Meta-Score
(Long *et al.*, 2021)	The benefits of defibrillator in heart failure patients with cardiac resynchronization therapy: A meta-analysis
(Vakil *et al.*, 2016)	Implantable Cardioverter-Defibrillator Use in Patients With Left Ventricular Assist Devices: A Systematic Review and Meta-Analysis

(Rorris *et al.*, 2021)	Implantable cardioverter defibrillators in left ventricular assist device patients: A systematic review and meta-analysis
(Carmo *et al.*, 2018)	Implantable cardioverter-defibrillator in Chagas heart disease: A systematic review and meta-analysis of observational studies
(Shurrab *et al.*, 2018)	Outcomes of ICDs and CRTs in patients with chronic kidney disease: a meta-analysis of 21,000 patients
(Sze and Daubert, 2015)	Why the Authors Use Cardiac Resynchronization Therapy with Defibrillators
(Barra *et al.*, 2015)	Importance of Implantable Cardioverter-Defibrillator Back-Up in Cardiac Resynchronization Therapy Recipients: A Systematic Review and Meta-Analysis
(Zhang, Zhou and Yu, 2015)	Incidence, definition, diagnosis, and management of the cardiac resynchronization therapy nonresponder
(Cleland *et al.*, 2022)	The effect of cardiac resynchronization without a defibrillator on morbidity and mortality: an individual patient data meta-analysis of COMPANION and CARE-HF
(Friedman *et al.*, 2017)	New York Heart Association class and the survival benefit from primary prevention implantable cardioverter defibrillators: A pooled analysis of 4 randomized controlled trials
(Yuyun *et al.*, 2021)	Risk of ventricular arrhythmia in cardiac resynchronization therapy responders and super-responders: a systematic review and meta-analysis
(Sipahi *et al.*, 2015)	Impact of QRS duration on survival benefit with prophylactic implantable cardioverter-defibrillators: a meta-analysis of randomized controlled trials
(Kanitsoraphan *et al.*, 2019)	Baseline fragmented QRS is associated with increased all-cause mortality in heart failure with reduced ejection fraction: A systematic review and meta-analysis
(Wasiak *et al.*, 2023)	An implantable cardioverter-defibrillator for primary prevention in non-ischemic cardiomyopathy: A systematic review and meta-analysis
(Hindricks *et al.*, 2017)	Daily remote monitoring of implantable cardioverter-defibrillators: insights from the pooled patient-level data from three randomized controlled trials (IN-TIME, ECOST, TRUST)
(Tseng *et al.*, 2019)	Efficacy of Pharmacologic and Cardiac Implantable Electronic Device Therapies in Patients With Heart Failure and Reduced Ejection Fraction: A Systematic Review and Network Meta-Analysis
(Elkaryoni *et al.*, 2019)	Implantable cardioverter-defibrillators and survival in advanced heart failure patients with continuous-flow left ventricular assist devices: a systematic review and meta-analysis
(Naka *et al.*, 2019)	Association between atrial fibrillation and patient-important outcomes in heart failure patients with implantable cardioverter-defibrillators: a systematic review and meta-analysis
(Mustafa *et al.*, 2018)	Atrial Fibrillation Is Associated With Higher Overall Mortality in Patients With Implantable Cardioverter-Defibrillator: A Systematic Review and Meta-Analysis
(Yue *et al.*, 2020)	Prognostic Value of Late Gadolinium Enhancement in Predicting Life-Threatening Arrhythmias in Heart Failure Patients With Implantable Cardioverter-Defibrillators: A Systematic Review and Meta-Analysis
(Batchelor *et al.*, 2023)	Meta-Analysis on Drug and Device Therapy of New York Heart Association Functional Class IV Heart Failure With Reduced Ejection Fraction
(Barra *et al.*, 2018)	Cause-of-death analysis in patients with cardiac resynchronization therapy with or without a defibrillator: a systematic review and proportional meta-analysis
(Nikolaidou *et al.*, 2018)	Postmortem ICD interrogation in mode of death classification

(Agrawal *et al.*, 2016)	The role of implantable cardioverter-defibrillators in patients with continuous flow left ventricular assist devices - A meta-analysis
(Kang *et al.*, 2015)	Cardiac resynchronization therapy and QRS duration: systematic review, meta-analysis, and meta-regression
(Bergau, Seegers and Zabel, 2014)	Sex differences in ICD benefit
(Israel and Manegold, 2014)	[Electrical storm: definition, prevalence, causes and prognostic implications]
(Earley *et al.*, 2014)	Effectiveness of implantable cardioverter defibrillators for primary prevention of sudden cardiac death in subgroups a systematic review
(Alba *et al.*, 2013)	Predictors of mortality in patients with an implantable cardiac defibrillator: a systematic review and meta-analysis
(Cleland *et al.*, 2013)	An individual patient meta-analysis of five randomized trials assessing the effects of cardiac resynchronization therapy on morbidity and mortality in patients with symptomatic heart failure
(STAVRAKIS, PATEL and REYNOLDS, 2013)	Defibrillation threshold testing does not predict clinical outcomes during long-term follow-up: a meta-analysis
(Tu *et al.*, 2013)	[A systematic review and meta-analysis on efficacy and safety of cardiac resynchronization therapy alone or in combination with implantable cardioversion defibrillation in patients with mild to severe heart failure]
(S. Chen *et al.*, 2013)	The efficacy and safety of cardiac resynchronization therapy combined with implantable cardioverter defibrillator for heart failure: a meta-analysis of 5674 patients
(Hess *et al.*, 2013)	Survival benefit of primary prevention implantable cardioverter-defibrillator therapy after myocardial infarction: does time to implant matter? A meta-analysis using patient-level data from 4 clinical trials
(Betts *et al.*, 2013)	Absolute risk reduction in total mortality with implantable cardioverter defibrillators: analysis of primary and secondary prevention trial data to aid risk/benefit analysis
(CHATTERJEE *et al.*, 2013)	Pharmacologic rate versus rhythm-control strategies in atrial fibrillation: an updated comprehensive review and meta-analysis
(Korantzopoulos *et al.*, 2009)	Implantable cardioverter defibrillator therapy in chronic kidney disease: a meta-analysis
(Rossi *et al.*, 2008)	The current role of cardiac resynchronization therapy in reducing mortality and hospitalization in heart failure patients: a meta-analysis from clinical trials
(Bradley *et al.*, 2003)	Cardiac resynchronization and death from progressive heart failure: a meta-analysis of randomized controlled trials

Search strategy for finding drug or non-digital biomarker-based meta-analyses in heart failure patients for mortality

Search strategy
((((("heart failure"[Title/Abstract] OR "heart failure"[MeSH Terms]) AND ("all-cause mortality"[Title/Abstract] OR "mortality"[MeSH Terms])) NOT "defibrillators, implantable"[MeSH Terms]) AND ("randomized controlled trial"[Publication Type] OR "controlled clinical trial"[Publication Type] OR "randomized"[Title/Abstract] OR "randomised"[Title/Abstract] OR "placebo"[Title/Abstract] OR "drug therapy"[MeSH

Subheading] OR "randomly"[Title/Abstract] OR "trial"[Title/Abstract] OR "groups"[Title/Abstract])) AND (meta-analysis[Filter])

The list of included drug or non-digital biomarker-based systematic reviews in heart failure patients in mortality

Study	Title
(Dinicolantonio *et al.*, 2013)	Meta-Analysis of Carvedilol Versus Beta 1 Selective Beta-Blockers (Atenolol, Bisoprolol, Metoprolol, and Nebivolol)
(Ng and Yap, 2018)	Continuous infusion vs. intermittent bolus injection of furosemide in acute decompensated heart failure: systematic review and meta-analysis of randomised controlled trials
(Butler *et al.*, 2020)	Efficacy and safety of SGLT2 inhibitors in heart failure: systematic review and meta-analysis
(Nielsen *et al.*, 2020)	Beneficial and harmful effects of sacubitril/valsartan in patients with heart failure: a systematic review of randomized clinical trials with meta-analysis and trial sequential analysis
(Anker *et al.*, 2018)	Effects of ferric carboxymaltose on hospitalisations and mortality rates in iron-deficient heart failure patients: an individual patient data meta-analysis
(Alskaf, Tridente and Al-Mohammad, 2016)	Tolvaptan for Heart Failure, Systematic Review and Meta-Analysis of Trials
(Berbenetz and Mrkobrada, 2016)	Mineralocorticoid receptor antagonists for heart failure: systematic review and meta-analysis
(Abdulla, Køber, *et al.*, 2006)	Effect of beta-blocker therapy on functional status in patients with heart failure — A meta-analysis
(Jin *et al.*, 2010)	A meta-analysis of erythropoiesis-stimulating agents in anaemic patients with chronic heart failure
(Gao *et al.*, 2011)	Trimetazidine: a meta-analysis of randomized controlled trials in heart failure
(De Vecchis *et al.*, 2015)	Hypertonic saline plus i.v. furosemide improve renal safety profile and clinical outcomes in acute decompensated heart failure
(Xiong *et al.*, 2015)	The short-term and long-term effects of tolvaptan in patients with heart failure: a meta-analysis of randomized controlled trials
(Qian *et al.*, 2016)	The Efficacy and Safety of Iron Supplementation in Patients With Heart Failure and Iron Deficiency: A Systematic Review and Meta-analysis
(Nistor *et al.*, 2015)	Vasopressin receptor antagonists for the treatment of heart failure: a systematic review and meta-analysis of randomized controlled trials
(Zheng *et al.*, 2018)	Drug treatment effects on outcomes in heart failure with preserved ejection fraction: a systematic review and meta-analysis
(Martin *et al.*, 2018)	Beta-blockers and inhibitors of the renin-angiotensin aldosterone system for chronic heart failure with preserved ejection fraction (Review)

(Wang, Xiong and Cai, 2017)	Effects of Tolvaptan in patients with acute heart failure: a systematic review and meta-analysis
(Song *et al.*, 2017)	Efficacy and Safety of L-Carnitine Treatment for Chronic Heart Failure: A Meta-Analysis of Randomized Controlled Trials
(Huang *et al.*, 2018)	Use of tolvaptan vs. furosemide in older patients with heart failure Meta-analysis of randomized controlled trials
(M. S. Khan *et al.*, 2017)	Dose of Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers and Outcomes in Heart Failure A Meta-Analysis

Search strategy for finding wearable activity tracker-based meta-analyses

Search strategy	Number of identified studies
("activity tracker"[Title/Abstract] OR "fitness trackers"[MeSH Terms] OR "fitbit"[Title/Abstract] OR "Garmin"[Title/Abstract] OR "Apple Watch"[Title/Abstract] OR "Jawbone"[Title/Abstract] OR "pedometer"[Title/Abstract] OR "Accelerometer"[Title/Abstract] OR "wearable"[Title/Abstract] OR "step count"[Title/Abstract] OR "wristband"[Title/Abstract] OR "app"[Title/Abstract]) AND ("weight"[Title/Abstract] OR "Body Weight"[MeSH Terms]) AND "meta analysis"[Publication Type]	44

The list of included wearable activity tracker-based systematic reviews in changing weight

Study	Title
(Wong *et al.*, 2022)	Wearable technology-delivered lifestyle intervention amongst adults with overweight and obese: A systematic review and meta-regression
(Yen, Jin and Chiu, 2023)	Smartphone app-based interventions targeting physical activity for weight management: A meta-analysis of randomized controlled trials
(Fakih El Khoury *et al.*, 2019)	The Effects of Dietary Mobile Apps on Nutritional Outcomes in Adults with Chronic Diseases: A Systematic Review and Meta-Analysis
(Dehghan Ghahfarokhi *et al.*, 2022)	The Effect of Wearable and Smartphone Applications on Physical Activity, Quality of Life, and Cardiovascular Health Outcomes in Overweight/Obese Adults: A Systematic Review and Meta-analysis of Randomized Controlled Trials
(Baskerville *et al.*, 2017)	Impact of accelerometer and pedometer use on physical activity and glycaemic control in people with Type 2 diabetes: a systematic review and meta-analysis
(Cai *et al.*, 2016)	Systematic Review or Meta-analysis Pedometer intervention and weight loss in overweight and obese adults with Type 2 diabetes: a meta-analysis
(Hsin Yen Yen and Chiu, 2019)	The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials
(Cai *et al.*, 2020)	Mobile Application Interventions and Weight Loss in Type 2 Diabetes: A Meta-Analysis

(Ringeval *et al.*, 2020)	Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis
(W. Wang *et al.*, 2022)	The Effectiveness of Wearable Devices as Physical Activity Interventions for Preventing and Treating Obesity in Children and Adolescents: Systematic Review and Meta-analysis
(Tang *et al.*, 2020)	Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials
(Chew *et al.*, 2022)	Sustainability of Weight Loss Through Smartphone Apps: Systematic Review and Meta-analysis on Anthropometric, Metabolic, and Dietary Outcomes
(Antoun *et al.*, 2022)	The Effectiveness of Combining Nonmobile Interventions With the Use of Smartphone Apps With Various Features for Weight Loss: Systematic Review and Meta-analysis
(Cui *et al.*, 2016)	T2DM Self-Management via Smartphone Applications: A Systematic Review and Meta-Analysis

The list of excluded wearable activity tracker-based systematic reviews

Study	Title
(Bourke *et al.*, 2023)	Adherence to the World Health Organization's physical activity recommendation in preschool-aged children: a systematic review and meta-analysis of accelerometer studies
(Musgrave *et al.*, 2023)	Addressing Preconception Behavior Change Through Mobile Phone Apps: Systematic Review and Meta-analysis
(Chew *et al.*, 2023)	Effectiveness of Combined Health Coaching and Self-Monitoring Apps on Weight-Related Outcomes in People With Overweight and Obesity: Systematic Review and Meta-analysis
(Peng *et al.*, 2023)	Effectiveness of Wearable Activity Monitors on Metabolic Outcomes in Patients With Type 2 Diabetes: A Systematic Review and Meta-Analysis
(Mamalaki *et al.*, 2022)	The effectiveness of technology-based interventions for weight loss maintenance: A systematic review of randomized controlled trials with meta-analysis
(Ang *et al.*, 2021)	Efficacy of Interventions That Incorporate Mobile Apps in Facilitating Weight Loss and Health Behavior Change in the Asian Population: Systematic Review and Meta-analysis
(Petkovic *et al.*, 2021)	Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population
(McDonough, Su and Gao, 2021)	Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis
(Saeteaw *et al.*, 2021)	Efficacy and safety of pharmacological cachexia interventions: systematic review and network meta-analysis
(Kamei *et al.*, 2022)	The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: A systematic review and meta-analysis
(Islam *et al.*, 2020)	Use of Mobile Phone App Interventions to Promote Weight Loss: Meta-Analysis
(Wiersma *et al.*, 2020)	Unravelling the association between accelerometer-derived physical activity and adiposity among preschool children: A systematic review and meta-analyses
(Farooq *et al.*, 2020)	Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: A systematic review and meta-analysis

(Mahajan *et al.*, 2020)	Complex interaction of obesity, intentional weight loss and heart failure: a systematic review and meta-analysis
(Georgarakis, Wolf and Riener, 2019)	Simplifying Exosuits: Kinematic Couplings in the Upper Extremity during Daily Living Tasks
(Villinger *et al.*, 2019)	The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis
(Matsui *et al.*, 2019)	Switching to antipsychotic monotherapy vs. staying on antipsychotic polypharmacy in schizophrenia: A systematic review and meta-analysis
(Chan and Chen, 2019)	Effects of Social Media and Mobile Health Apps on Pregnancy Care: Meta-Analysis
(Love, Adams and Sluijs, 2019)	Are school-based physical activity interventions effective and equitable? A meta-analysis of cluster randomized controlled trials with accelerometer-assessed activity
(Dunn, White and Green, 2018)	A model to estimate seabird field metabolic rates
(Silva *et al.*, 2018)	What is the effect of diet and/or exercise interventions on behavioural compensation in non-exercise physical activity and related energy expenditure of free-living adults? A systematic review
(Lunde *et al.*, 2018)	The Effectiveness of Smartphone Apps for Lifestyle Improvement in Noncommunicable Diseases: Systematic Review and Meta-Analyses
(Schock, Neher and Safranek, 2017)	Clinical Inquiry: Do pedometers increase activity and improve health outcomes?
(Borde *et al.*, 2017)	Methodological considerations and impact of school-based interventions on objectively measured physical activity in adolescents: a systematic review and meta-analysis
(Herring *et al.*, 2016)	Changes in physical activity behaviour and physical function after bariatric surgery: a systematic review and meta-analysis
(Flores Mateo *et al.*, 2015)	Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis
(Jeon and Park, 2015)	Nursing Intervention using smartphone technologies; a systematic review and meta-analysis
(Nascimento *et al.*, 2014)	The effect of physical exercise strategies on weight loss in postpartum women: a systematic review and meta-analysis
(Metcalf, Henley and Wilkin, 2012)	Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54)
(Richardson *et al.*, 2008)	A meta-analysis of pedometer-based walking interventions and weight loss

The list of the included drug or non-wearable activity trackers systematic reviews in changing weight

Study	Title

(Maharlouei *et al.*, 2019)	The effects of ginger intake on weight loss and metabolic profiles among overweight and obese subjects: A systematic review and meta-analysis of randomized controlled trials
(Payab *et al.*, 2020)	Effect of the herbal medicines in obesity and metabolic syndrome: A systematic review and meta-analysis of clinical trials
(Yang *et al.*, 2022)	The effects of low-fat, high-carbohydrate diets vs. low-carbohydrate, high-fat diets on weight, blood pressure, serum liquids and blood glucose: a systematic review and meta-analysis
(De Menezes *et al.*, 2019)	Influence of Paleolithic diet on anthropometric markers in chronic diseases: systematic review and meta-analysis
(Min *et al.*, 2017)	Comparison between SGLT2 inhibitors and DPP4 inhibitors added to insulin therapy in type 2 diabetes: a systematic review with indirect comparison meta-analysis
(Paravattil, Wilby and Turgeon, 2016)	Topiramate monotherapy for weight reduction in patients with type 2 diabetes mellitus: A systematic review and meta-analysis
(Meng *et al.*, 2017)	Efficacy of low carbohydrate diet for type 2 diabetes mellitus management: A systematic review and meta-analysis of randomized controlled trials
(Avgerinos *et al.*, 2020)	Oral semaglutide for type 2 diabetes: A systematic review and meta-analysis
(Duan *et al.*, 2020)	Effects of Vitamin D Supplementation on General and Central Obesity: Results from 20 Randomized Controlled Trials Involving Apparently Healthy Populations
(C. Zhou *et al.*, 2022)	Ketogenic Diet Benefits to Weight Loss, Glycemic Control, and Lipid Profiles in Overweight Patients with Type 2 Diabetes Mellitus: A Meta-Analysis of Randomized Controlled Trails
(Guo *et al.*, 2022)	The Antiobesity Effect and Safety of GLP-1 Receptor Agonist in Overweight/Obese Patients Without Diabetes: A Systematic Review and Meta-Analysis
(Zhong *et al.*, 2022)	Efficacy and safety of once-weekly semaglutide in adults with overweight or obesity: a meta-analysis

Appendix III. The methodological quality assessment of included SRs

The methodological quality assessment of digital biomarkers-based systematic reviews using AMSTAR-2 tool

Studies	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Overall point	AMSTAR grade
(Wong *et al.*, 2022)																	13	L
(Yen, Jin and Chiu, 2023)																	10	CL
(Fakih El Khoury *et al.*, 2019)																	11	L
(Dehghan Ghahfarokhi *et al.*, 2022)																	10	CL
(Baskerville *et al.*, 2017)																	12	CL
(Cai *et al.*, 2016)																	10	L
(Hsin Yen Yen and Chiu, 2019)																	10	CL
(Cai *et al.*, 2020)																	8	CL
(Ringeval *et al.*, 2020)																	12	CL
(W. Wang *et al.*, 2022)																	9	CL
(Tang *et al.*, 2020)																	10	CL
(Chew *et al.*, 2022)																	12	L

(Antoun *et al.*, 2022)	10	CL
(Cui *et al.*, 2016)	11	CL
(Alba, Forouta n, *et al.*, 2018)	12	CL
(Huang *et al.*, 2010)	7	CL
(Bertold i *et al.*, 2011)	7	CL
(Kang *et al.*, 2015)	11	CL
(Shah *et al.*, 2014)	10	CL
(Shaojie Chen *et al.*, 2013)	6	CL
(Xing *et al.*, 2017)	10	CL
(Sun *et al.*, 2016)	7	CL
(Shun-Shin *et al.*, 2017)	11	CL
(Gama *et al.*, 2020)	9	CL
(Akel and Lafferty , 2017)	6	CL
(Barakat *et al.*, 2017)	12	M
(Alotaib i *et al.*, 2020)	12	L
(Al-Khatib *et al.*, 2017)	4	CL
(Miller *et al.*, 2015)	8	CL
(Siddiqu i *et al.*, 2018)	11	CL

(Wolff *et al.*, 2017)																	8	CL
(El Moheb *et al.*, 2018)																	16	H
(Ghanbari, 2009)																	11	CL
(Abdulla, Haarbo, *et al.*, 2006)																	7	CL

Q1: PICO, Q2: Protocol, Q3: Selection of study design, Q4: Comprehensive literature search, Q5: Study selection in duplicate, Q6: Data extraction in duplicate, Q7: List of excluded studies, Q8: Describe the included studies in adequate detail, Q9: A satisfactory technique for assessing the risk of bias, Q10: Sources of funding for the studies included, Q11: Appropriate methods for statistical combination of results, Q12: Potential impact of RoB in individual studies on the results of the meta-analysis, Q13: RoB in individual studies when interpreting/ discussing the results of the review, Q14: A satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review, Q15: An adequate investigation of publication bias, Q16: Potential sources of conflict of interest, ROB: Risk of bias, H: High methodological quality, M: Moderate methodological quality, L: Low methodological quality, CL: Critically low methodological quality, Green color: Yes, Red color: No, Yellow color: Partially yes.

The methodological quality assessment of non-digital biomarkers or drug-based systematic reviews using AMSTAR-2 tool

Studies	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Overall point	AMSTAR grade
(Maharlouei *et al.*, 2019)																	9	CL
(Payab *et al.*, 2020)																	10	L
(Yang *et al.*, 2022)																	10	CL
(De Menezes *et al.*, 2019)																	6	CL
(Min *et al.*, 2017)																	9	CL
(Paravattil, Wilby and Turgeon, 2016)																	9	CL
(Meng *et al.*, 2017)																	9	CL

(Avgerinos *et al.*, 2020)	10	CL
(Duan *et al.*, 2020)	11	L
(C. Zhou *et al.*, 2022)	8	CL
(Guo *et al.*, 2022)	11	L
(Zhong *et al.*, 2022)	11	M
(Arastu *et al.*, 2022)	10	CL
(Toledo *et al.*, 2023)	12	L
(Dinicolantonio *et al.*, 2013)	8	CL
(Ng and Yap, 2018)	11	L
(Butler *et al.*, 2020)	11	CL
(Nielsen *et al.*, 2020)	12	CL
(Anker *et al.*, 2018)	10	CL
(Alskaf, Tridente and Al-Mohammad, 2016)	11	L
(Berbenetz and Mrkobrada, 2016)	9	CL
(Abdulla, Køber, *et al.*, 2006)	9	CL
(Jin *et al.*, 2010)	8	CL
(Gao *et al.*, 2011)	12	CL
(De Vecchis *et al.*, 2015)	9	CL
(Xiong *et al.*, 2015)	9	CL
(Qian *et al.*, 2016)	6	CL
(Nistor *et al.*, 2015)	12	CL
(Zheng *et al.*, 2018)	8	CL
(Martin *et al.*, 2018)	16	H

(Wang, Xiong and Cai, 2017)	8	CL	
(Song *et al.*, 2017)	9	CL	
(Huang *et al.*, 2018)	11	L	
(M. S. Khan *et al.*, 2017)	8	CL	

Q1: PICO, Q2: Protocol, Q3: Selection of study design, Q4: Comprehensive literature search, Q5: Study selection in duplicate, Q6: Data extraction in duplicate, Q7: List of excluded studies, Q8: Describe the included studies in adequate detail, Q9: A satisfactory technique for assessing the risk of bias, Q10: Sources of funding for the studies included, Q11: Appropriate methods for statistical combination of results, Q12: Potential impact of RoB in individual studies on the results of the meta-analysis, Q13: RoB in individual studies when interpreting/ discussing the results of the review, Q14: A satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review, Q15: An adequate investigation of publication bias, Q16: Potential sources of conflict of interest, ROB: Risk of bias, H: High methodological quality, M: Moderate methodological quality, L: Low methodological quality, CL: Critically low methodological quality, Green color: Yes, Red color: No, Yellow color: Partially yes.

Appendix IV. The quality evidence assessment results

The quality evidence assessment results of digital biomarker-based meta-analyses using the GRADE tool.

Study	Outcome	Effect size (95% CIs)	Consistency	Imprecision	ROB	PB	Indirectness	Quality of evidence	Comment
(Alba, Foroutan, *et al.*, 2018)	All-cause mortality	0.84 (0.73, 0.96)/ RR	0	0	0	0	0	High	
(Huang *et al.*, 2010)	All-cause mortality	0.55 (0.4, 0.76)/ OR	0	0	-1	0	0	Moderate	43% of the included studies have low risk of bias
(Bertoldi *et al.*, 2011)	All-cause mortality	0.83 (0.72, 0.96) /RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Kang *et al.*, 2015)	All-cause mortality	0.81 (0.68, 0.97)/ OR	0	0	-1	0	0	Moderate	26% of the included studies have low risk of bias
(Shah *et al.*, 2014)	All-cause mortality	1.66 (1.096, 1.515)/RR	-1	-1	0	0	0	Moderate	Heterogeneity not reported The optimum information size is 1858
(Shaojie Chen *et al.*, 2013)	Mortality	0.8 (0.67, 0.95)/ OR	0	0	-1	0	0	Moderate	50% of the included studies have low risk of bias
(Xing *et al.*, 2017)	All-cause mortality	0.83 (0.71, 0.97)/ RR	0	0	-1	0	0	Moderate	40% of the included studies have low risk of bias
(Sun *et al.*, 2016)	All-cause mortality	0.78 (0.63, 0.96)/ OR	0	0	0	0	0	High	
(Shun-Shin *et al.*, 2017)	All-cause mortality	0.76 (0.64, 0.9)/ HR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Gama *et al.*, 2020)	All-cause mortality	0.85 (0.78, 0.94)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias

(Akel and Lafferty, 2017)	All-cause mortality	0.8 (0.76, 0.96)/ HR	0	0	-1	0	0	Moderate	Risk of bias of the included studies is not reported
(Barakat *et al.*, 2017)	All-cause mortality	0.79 (0.64, 0.93)/ HR	0	0	0	0	0	High	
(Alotaibi *et al.*, 2020)	All-cause mortality	0.88 (0.69, 1.11)RR	0	0	-1	0	0	Moderate	43% of the included studies have low risk of bias
(Al-Khatib *et al.*, 2017)	All-cause mortality	0.75 (0.61, 0.93)/ HR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Miller *et al.*, 2015)	Mortality	0.63 (0.52, 0.75)/ RR	-1	0	-1	0	0	Moderate	Heterogeneity not reported None of the included studies has low risk of bias
(Siddiqui *et al.*, 2018)	All-cause mortality	0.74 (0.62, 0.9)/ OR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Wolff *et al.*, 2017)	All-cause mortality	0.77 (0.64, 0.93)/ OR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(El Moheb *et al.*, 2018)	All-cause mortality	0.78 (0.66, 0.92)/ HR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Ghanbari, 2009)	Mortality	0.78 (0.7, 0.87)/ HR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Abdulla, Haarbo, *et al.*, 2006)	All-cause mortality	0.75 (0.59, 0.96)/ OR	-1	0	-1	0	0	Moderate	Heterogeneity not reported None of the included studies has low risk of bias
(Wong *et al.*, 2022)	Weight	-0.95 (-3.08, 1.18)/ MD	0	-1	-1	0	0	Moderate	The optimum sample size is 27480 38% of the included studies have low risk of bias
(Yen, Jin and Chiu, 2023)	Weight	-0.434 (-0.684, -0.184)/ Hg	-1	0	-1	0	0	Moderate	Heterogeneity is 78.04% 30% of the included studies have low risk of bias

(Fakih El Khoury *et al.*, 2019)	Weight	-2.45 (-3.33, -1.58)/ WMD	-1	0	-1	0	0	Moderate	Heterogeneity is 96.2% 9% of the included studies have low risk of bias
(Dehghan Ghahfarokhi *et al.*, 2022)	Weight	-1.61 (-2.82, -0.39)/ MD	0	-1	-1	0	0	Moderate	The optimal sample size is 2476 23% of the included studies have low risk of bias
(Baskerville *et al.*, 2017)	Body mass index	0.06 (-0.19, 0.32)/ SMD	0	-1	-1	0	0	Moderate	The optimal sample size is 6872 None of the included studies has low risk of bias
(Cai *et al.*, 2016)	Weight	-0.65 (-1.12, -0.17)/ WMD	0	-1	-1	0	0	Moderate	The optimal information size is 1466 28% of the included studies have low risk of bias
(Hsin Yen Yen and Chiu, 2019)	Weight	-0.594 (-0.842, -0.346)/ Hg	-1	0	-1	0	0	Moderate	Heterogeneity is 86.07% None of the included studies has low risk of bias
(Cai *et al.*, 2020)	Weight	-0.84 (-1.151, -0.17)/ WMD	0	-1	-1	0	0	Moderate	The optimal information is 1720 11% of the included studies have low risk of bias
(Ringeval *et al.*, 2020)	Weight	-1.48 (-2.81, -0.14)/ MD	0	-1	-1	0	0	Moderate	The optimal information size is 5050 None of the included studies has low risk of bias
(W. Wang *et al.*, 2022)	Weight	-1.08 (-2.16, 0)/ MD	0	-1	-1	0	0	Moderate	The optimal information size is 3056 None of the included studies has low risk of bias

(Tang *et al.*, 2020)	Weight	0.133 (-0.336, 0.603)/ SMD	0	-1	-1	0	0	Moderate	The optimal information size is 1400 None of the included studies has low risk of bias
(Chew *et al.*, 2022)	Weight	-1.15 (-3.02, 0.72)/ MD	-1	-1	-1	0	0	Low	Heterogeneity is 91% The optimal information size is 15458 None of the included studies has low risk of bias
(Antoun *et al.*, 2022)	Weight	-1.95 (-2.09, -1.81)/ MD	-1	0	-1	0	0	Moderate	Heterogeneity is 81% 17% of the included studies have low risk of bias
(Cui *et al.*, 2016)	Weight	-0.84 (-2.04, 0.36)/ MD	0	-1	-1	0	0	Moderate	The optimal information size is 5050 None of the included studies has low risk of bias

CIs: Confidence intervals, ROB: risk of bias, PB: publication bias, RR: Risk ratio, OR: Odds ratio, HR: Hazard ratio, MD: Mean difference, Hg: Hedges' g, WMD: Weighted mean difference.

The quality of evidence of non-digital biomarker or drug-based meta-analyses using the GRADE tool

Study	Outcome	Effect size (95% CIs)	Consistency	Imprecision	ROB	PB	Indirectness	Quality of evidence	Comment
(Dinicolantonio *et al.*, 2013)	All-cause mortality	0.85 (0.78, 0.93)/ RR	0	0	-1	0	0	Moderate	12.5% of the included studies have low risk of bias
(Ng and Yap, 2018)	All-cause mortality	1.65 (0.93, 2.91)/ OR	0	-1	-1	0	0	Moderate	The optimal information size is 1132

									25% of the included studies have low risk of bias
(Butler *et al.*, 2020)	All-cause mortality	1.02 (0.79, 1.3)/ HR	0	0	0	0	0	High	
(Nielsen *et al.*, 2020)	All-cause mortality	0.86 (0.79, 0.94)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Anker *et al.*, 2018)	Cardiovascular mortality	0.59 (0.4, 0.88)/ Rate ratio	0	0	-1	0	0	Moderate	Risk of bias of the included studies is not reported
(Alskaf, Tridente and Al-Mohammad, 2016)	Mortality	0.81 (0.51, 1.3)/ OR	0	0	-1	0	0	Moderate	66% of the included studies have low risk of bias
(Berbenetz and Mrkobrada, 2016)	All-cause mortality	0.83 (0.77, 0.88)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Abdulla, Køber, *et al.*, 2006)	All-cause mortality	0.69 (0.59, 0.82)/ OR	0	0	-1	0	0	Moderate	56% of the included studies have low risk of bias
(Jin *et al.*, 2010)	All-cause mortality	0.71 (0.41, 1.24)/ RR	0	-1	-1	0	0	Moderate	The optimal information size is 2496 16% of the included studies have low risk of bias
(Gao *et al.*, 2011)	All-cause mortality	0.29 (0.17, 0.49)/ RR	0	-1	-1	0	0	Moderate	The optimal information size is 572 66% of the included studies have low risk of bias
(De Vecchis *et al.*, 2015)	All-cause mortality	0.57 (0.44, 0.74)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Xiong *et al.*, 2015)	All-cause mortality	0.96 (0.87, 1.06)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias

(Qian *et al.*, 2016)	All-cause mortality	0.81 (0.42, 1.57)/ OR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Nistor *et al.*, 2015)	All-cause mortality	0.98 (0.88, 1.08)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Zheng *et al.*, 2018)	All-cause mortality	0.96 (0.9, 1.03)/ RR	0	0	-1	0	0	Moderate	37% of the included studies have low risk of bias
(Martin *et al.*, 2018)	All-cause mortality	0.99 (0.71, 1.38) RR	0	0	-1	0	0	Moderate	40% of the included studies has low risk of bias
(Wang, Xiong and Cai, 2017)	All-cause mortality	0.98 (0.68, 1.43)/ RR	0	-1	-1	0	0	Moderate	The optimal information size is 2140 None of the included studies has low risk of bias
(Song *et al.*, 2017)	All-cause mortality	0.48 (0.21, 1.06)/ OR	0	-1	-1	0	0	Moderate	The optimal information size is 1118 None of the included studies has low risk of bias
(Huang *et al.*, 2018)	All-cause mortality	0.73 (0.36, 1.47)/ RR	0	-1	-1	0	0	Moderate	The optimal information size is 2874 50% of the included studies have low risk of bias
(M. S. Khan *et al.*, 2017)	All-cause mortality	0.94 (0.89, 1)/ RR	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Maharlouei *et al.*, 2019)	Weight	-0.66 (-1.33, -0.01)/ SMD	-1	-1	-1	0	0	Low	Heterogeneity is 76.9% None of the included studies has low risk of bias
(Payab *et al.*, 2020)	Weight	-0.75 (-1.19,	-1	0	-1	0	0	Moderate	Heterogeneity is 94.3%

		-0.32)/ SMD							18% of the included studies have low risk of bias
(Yang *et al.*, 2022)	Weight	-1.01 (-1.96, -0.77)/ SMD	-1	0	-1	0	0	Moderate	Heterogeneity is 95% 42% of the included studies have low risk of bias
(De Menezes *et al.*, 2019)	Weight	-3.52 (-5.26, -1.79)/ MD	0	-1	-1	0	0	Moderate	The optimum information size 652 None of the included studies has low risk of bias
(Min *et al.*, 2017)	Weight	-0.04 (-0.25, 0.16)/ WMD	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Paravattil, Wilby and Turgeon, 2016)	Weight	-3.41 (-3.79, -3.04)/ MD	0	0	-1	0	0	Moderate	None of the included studies has low risk of bias
(Meng *et al.*, 2017)	Weight	-0.94 (-1.92, -0.05)/ WMD	0	-1	-1	0	0	Moderate	The optimum information size 3878 62.5% of the included studies have low risk of bias
(Avgerinos *et al.*, 2020)	Weight	-2.99 (-3.69, -2.3)/ MD	-1	0	0	0	0	Moderate	Heterogeneity is 78%
(Duan *et al.*, 2020)	Body mass index	-0.09 (-0.19, 0.01)/ WMD	0	0	-1	0	0	Moderate	50% of the included studies have low risk of bias
(C. Zhou *et al.*, 2022)	Weight	-5.62 (-9.73, -1.51)/ MD	0	-1	-1	0	0	Moderate	The optimum information size 1860 None of the included studies has low risk of bias

(Guo *et al.*, 2022)	Weight	-5.39 (-6.82, -3.96)/ WMD	-1	0	-1	0	0	Moderate	Heterogeneity is 99.2% 17% of the included studies have low risk of bias
(Zhong *et al.*, 2022)	Weight	-11.9 (-13.24, -10.56)/ MD	0	0	0	0	0	High	
(Arastu *et al.*, 2022)	Weight	-11.62 (-13.03, -10.21)/ MD	0	0	0	0	0	High	
(Toledo *et al.*, 2023)	Body mass index	-0.1 (-0.27, 0.07)/ SMD	0	-1	-1	0	0	Moderate	Optimum information size is 3142 71% of the included studies have low risk of bias

CIs: Confidence intervals, ROB: risk of bias, PB: publication bias, RR: Risk ratio, OR: Odds ratio, HR: Hazard ratio, MD: Mean difference, Hg: Hedges' g, WMD: Weighted mean difference.

Appendix V. The characteristics of the included DBMs studies and the searches conducted to find non-DBMS or drug-based studies

DBM-based studies						Non-DBM or drug-based studies				
Study	Population	Intervention	Outcome	Number of missing studies (change in the magnitude of effect size %)	Search strategy to find drug-based meta-analyses	Study	Population	Intervention	Outcome	Number of missing studies (change in the magnitude of effect size %)

(Ringeval *et al.*, 2020)	Overweight, Sedentary, Arthritis, Cardiovascular risks, Diabetes, Cardiometabolic diseases, Chronic low back pain, Chronic obstructive pulmonary disease, Prediabetes, With low ankle brachial index, healthy subjects	Fitbit	steps	0 (0%)	((Overweight[Title/Abstract] OR Sedentary[Title/Abstract] OR Arthritis[Title/Abstract] OR Cardiovascular[Title/Abstract] OR Diabetes[Title/Abstract] OR Cardiometabolic[Title/Abstract] OR "Chronic low back pain"[Title/Abstract] OR "Chronic obstructive pulmonary disease"[Title/Abstract] OR "Prediabetes"[Title/Abstract] OR "low ankle brachial index"[Title/Abstract] OR "healthy subjects"[Title/Abstract]) NOT ("Fitness Trackers"[Mesh] OR "activity tracker"[title/abstract])) AND ("Body Weight"[Mesh] OR "Weight Loss"[Mesh] OR "Body Weight Changes"[Mesh]) limited to 2020	(Talenezhad *et al.*, 2020)	Numerous kind of populations	L-carnitine supplementation	BMI	0 (0%)
(Ringeval *et al.*, 2020)	Overweight, Sedentary, Arthritis, Cardiovascular risks, Diabetes, Cardiometab	Fitbit	MVPA	0 (0%)	((Overweight[Title/Abstract] OR Sedentary[Title/Abstract] OR Arthritis[Title/Abstract] OR Cardiovascular[Title/Abstr	(Talenezhad *et al.*, 2020)	Numerous kind of populations	L-carnitine supplementation	Weight	6 (32%)

	olic diseases, Chronic low back pain, Chronic obstructive pulmonary disease, Prediabetes, With low ankle brachial index, healthy subjects				act] OR Diabetes[Title/Abstract] OR Cardiometabolic[Title/Abstract] OR "Chronic low back pain"[Title/Abstract] OR "Chronic obstructive pulmonary disease"[Title/Abstract] OR "Prediabetes"[Title/Abstract] OR "low ankle brachial index"[Title/Abstract] OR "healthy subjects"[Title/Abstract]) NOT ("Fitness Trackers"[Mesh] OR "activity tracker"[title/abstract])) AND ("Body Weight"[Mesh] OR "Weight Loss"[Mesh] OR "Body Weight Changes"[Mesh]) limited to 2020					
(Ringeval *et al.*, 2020)	Overweight, Sedentary, Arthritis, Cardiovascular risks, Diabetes, Cardiometabolic diseases, Chronic low back pain, Chronic obstructive pulmonary	Fitbit	Weight	2 (14%)	((Overweight[Title/Abstract] OR Sedentary[Title/Abstract] OR Arthritis[Title/Abstract] OR Cardiovascular[Title/Abstract] OR Diabetes[Title/Abstract] OR Cardiometabolic[Title/Abstract] OR "Chronic low back pain"[Title/Abstract] OR "Chronic obstructive	(Yuan *et al.*, 2020)	Patients with T2DM	ketogenic diet	Weight	2 (12.6%)

	disease, Prediabetes, With low ankle brachial index, healthy subjects				pulmonary disease"[Title/Abstract] OR "Prediabetes"[Title/Abstract] OR "low ankle brachial index"[Title/Abstract] OR "healthy subjects"[Title/Abstract]) NOT ("Fitness Trackers"[Mesh] OR "activity tracker"[title/abstract])) AND ("Body Weight"[Mesh] OR "Weight Loss"[Mesh] OR "Body Weight Changes"[Mesh]) limited to 2021					
(Jang *et al.*, 2020)	Heart failure	Remote monitoring using digital devices	Detectio of atrial arrhythmia	4 (1.6%)	(("Heart Failure"[Mesh]) NOT ("Defibrillators, Implantable"[Mesh])) AND ("Arrhythmias, Cardiac"[Mesh])	(R. Wang *et al.*, 2022)	Heart failure	Sacubitril/Valsartan	Arrhythmias	1 (0.6%)
(Wang *et al.*, 2020)	cardiac arrest	Feedback device	ROSC	0 (0%)	(("cardiac arrest"[Title]) NOT (device[Title/Abstract])) AND (circulation[Title])	(An *et al.*, 2022)	Cardiopulmonary arrest	cumulative dose of adrenaline	ROSC	2 (1.2%)
(Tang *et al.*, 2020)	Healthy adults	Wearable Trackers	Physical activity	4 (68%)	(("healthy"[Title/abstract]) AND (weight[Title])) NOT ("Fitness Trackers"[Mesh] OR "activity tracker"[Title/Abstract])	(Abdollahi *et al.*, 2020)	General	Zinc Supplementation	Weight	0 (0%)

(Franssen *et al.*, 2020)	Type 2 diabetes mellitus	wearable activity tracker	Physical activity	0 (0%)	(("Diabetes Mellitus, Type 2"[Mesh] OR "type 2 diabetes"[Title]) NOT ("Fitness Trackers"[Mesh] OR "activity tracker"[Title/Abstract])) AND (weight[Title])	(Janani *et al.*, 2021)	Type 2 diabetes mellitus	Sitagliptin as Monotherapy and Add-On to Metformin	Weight	0 (0%)
(Alotaibi *et al.*, 2020)	Heart failure	Remote monitoring	Hospitalization	3 (10.8%)	(("Heart Failure"[Mesh] OR "heart failure"[Title/abstract]) NOT (monitoring[title/abstract] OR defibrillator[title/abstract])) AND (Hospitalization[Title/Abstract])	(Bamforth *et al.*, 2021)	Heart Failure, chronic obstructive pulmonary disease, and chronic kidney disease	Post-discharge interventions	Hospitalization	4 (4.7%)
(Hodkinson *et al.*, 2019)	Cardiometabolic Conditions	Accelerometer	Physical activity	5 (16.18 %)	((("Non-alcoholic Fatty Liver Disease"[Mesh] OR "non-alcoholic fatty liver"[title] OR "heart attack"[title/abstract] OR "Stroke"[Mesh] OR Stroke[title] OR "Diabetes Mellitus"[Mesh] OR Diabetes[Title] OR "Insulin Resistance"[Mesh] OR "Metabolic syndrome"[title]) AND ("Body Weight"[Mesh] OR weight[title])) AND ("meta analysis"[Publication Type])) AND (effect[title])	(Wu *et al.*, 2022)	People with diabetes	dual SGLT 1/2 inhibitor sotagliflozin	Weight	9 (12.65 %)

(Hodkinson *et al.*, 2019)	Cardiometabolic Conditions	Pedometer	Physical activity	0 (0%)	((("Non-alcoholic Fatty Liver Disease"[Mesh] OR "non-alcoholic fatty liver"[title] OR "heart attack"[title/abstract] OR "Stroke"[Mesh] OR Stroke[title] OR "Diabetes Mellitus"[Mesh] OR Diabetes[Title] OR "Insulin Resistance"[Mesh] OR "Metabolic syndrome"[title]) AND ("Body Weight"[Mesh] OR weight[title])) AND ("meta analysis"[Publication Type])) AND (effect[title])	(Wang, Zheng and Jin, 2021)	Acute decompensated heart failure	Ultrafiltration	Weight	1 (6.41%)
(Halawa, Enezate and Flaker, 2019)	Heart failure	Device monitoring	HF related readmission rate	2 (0.5%)	(Heart failure[Title/Abstract] AND "readmission"[Title/Abstract]) AND ("meta-analysis"[Publication Type])	(Campo *et al.*, 2017)	Myocardial infarction	Drugs targeting mitochondrial function	Hospital readmission	5 (16.36 %)
(Halawa, Enezate and Flaker, 2019)	Heart failure	Device monitoring	All-cause mortality	1 (0.52%)	((Heart failure[Title/Abstract]) AND ("Mortality"[Mesh] OR mortality[Title/Abstract])) AND ("meta-analysis"[Publication Type])	(McLellan *et al.*, 2020)	Heart failure	Natriuretic peptide-guided Treatment	All-cause mortality	2 (0.64%)
(Hsin Yen Yen and Chiu, 2019)	overweight	wearable technologies	Weight	2 (13.64 %)	((("Overweight"[Mesh] OR overweight[Title/Abstract]) OR ("Body Weight"[Mesh] OR weight[Title/Abstract])) NOT ("Fitness	(Moon *et al.*, 2021)	overweight	Liraglutide	Weight	1 (1.5%)

					Trackers"[Mesh] OR "Activity tracker"[Title/Abstract])) AND ("meta-analysis"[Publication Type])					
(Hsin Yen Yen and Chiu, 2019)	overweight	wearable technologies	BMI	0 (0%)	((("Overweight"[Mesh] OR overweight[Title/Abstract]) OR ("Body Weight"[Mesh] OR weight[Title/Abstract])) NOT ("Fitness Trackers"[Mesh] OR "Activity tracker"[Title/Abstract])) AND ("meta-analysis"[Publication Type])	(Moon *et al.*, 2021)	overweight	Liraglutide	BMI	4 (8.5%)
(Braakhuis, Berger and Bussmann, 2019)	All patients	wearable devices	Physical activity	1 (4.1%)	(("Body Weight"[Mesh] OR weight[Title/Abstract]) NOT ("Fitness Trackers"[Mesh] OR "Activity tracker"[Title/Abstract])) NOT ("meta-analysis"[Publication Type])	(De Menezes *et al.*, 2019)	Chronic diseases	Paleolithic diet	Weight	0 (0%)
(Kirk *et al.*, 2019)	Chronic Cardiometabolic Disease	Wearable Technology	steps	4 (26.77 %)	(((Cardiometabolic[Title/Abstract]) AND ("Body Weight"[Mesh] OR Weight[Title/Abstract])) NOT ("Fitness Trackers"[Mesh] OR "Activity tracker"[title/Abstract])) AND ("meta-analysis"[Publication Type])	(Min *et al.*, 2017)	type 2 diabetes	SGLT2 inhibitors and DPP4 Inhibitors	Weight	0 (0%)

(Kirk *et al.*, 2019)	Chronic Cardiometabolic Disease	Wearable Technology	MVPA	7 (72%)	(((Cardiometabolic[Title/Abstract]) AND ("Body Weight"[Mesh] OR Weight[Title/Abstract])) NOT ("Fitness Trackers"[Mesh] OR "Activity tracker"[title/Abstract])) AND ("meta-analysis"[Publication Type])	(Pan *et al.*, 2022)	type 2 diabetes	SGLT-2 inhibitors	Weight	3 (7.7%)
(Armstrong *et al.*, 2019)	Chronic obstructive pulmonary disease	pedometers	Physical activity	2 (13.3%)	((("Lung Diseases"[Mesh] OR "lung diseases, obstructive"[MeSH Terms] OR "Chronic obstructive pulmonary disease"[Title/Abstract]) NOT ("Fitness Trackers"[MeSH Terms] OR "activity tracker"[Title/Abstract] OR "wearable"[Title/Abstract] OR "Pedometer"[Title/Abstract])) AND ("Body Weight"[MeSH Terms] OR weight[title/abstract]) AND "meta analysis"[Publication Type]	(Shen *et al.*, 2018)	chronic obstructive pulmonary disease	Roflumilast	Weight	4 (1.5%)
(Kanitsorap han *et al.*, 2019)	Heart failure	fragmented QRS	Major arrhythmic events	4 (16%)	(("heart failure"[MeSH Terms] OR "heart failure"[Title/Abstract]) AND "Arrhythmias, Cardiac"[Mesh]) AND (meta-analysis[Filter])	(Sfairopoulos *et al.*, 2022)	Type 2 diabetes, heart failure, chronic kidney disease	sodium–glucose cotransporter-2 inhibitors	Ventricular arrhythmias	3 (0.14%)

(Kanitsorap han *et al.*, 2019)	Heart failure	fragmented QRS	All-cause mortality	5 (27.9%)	(("heart failure"[MeSH Terms] OR "heart failure"[Title/Abstract]) AND "Mortality"[Mesh]) AND (meta-analysis[Filter])	(Hansen *et al.*, 2020)	Heart failure	Pharmacologic al Heart Failure Treatment	All-cause mortality	0 (0%)

ROSC: Return of spontaneous circulation, BMI: Body mass index, T2DM: Type 2 diabetes mellitus, MVPA: Moderate to vigorous physical activity, HF: Heart failure.