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

Hossein Motahari Nezhad

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

Budapest, 2023

Hossein Motahari Nezhad

Corvinus University of Budapest

Department of Health Economics

Supervisor:

Dr. Zrubka Zsombor, MD, MBA, PhD.

Table of contents

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. 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). This thesis focus on a special class of digital medical devices: digital biomarkers (DBMs). *"Digital biomarkers are objective, measurable, physiological, and behavioural parameters collected using wearable, portable, implantable, or digestible digital devices"* (Babrak *et al.*, 2019).

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. The effect of these devices also depends on the change of human health behaviour or the decision of a health care professional.

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).

It is highly recommended by the Cochrane Handbook (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, 2022) 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, 2020), ICHI (International Classification of Health

Interventions) (World Health Organization, 2020), and ICF (International Classification of Functioning, Disability and Health) (WHO, 2017) tools. AMSTAR-2 (Shea *et al.*, 2017), is also one of the validated tools for researchers to evaluate the methodological quality of systematic reviews and meta-analyses. 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). 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).

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. In terms of rules, standards, etc., there are distinctions between digital and non-digital or

pharmaceutical products. No research has, to our knowledge, evaluated the methodological issues of digital biomarker-based studies and compared them to non-digital biomarkers. In this thesis, 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.

II. ESTABLISHING A CLASSIFICATION SYSTEM IN DIGITAL BIOMARKERS

II.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). 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). 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. We also assessed the usability of the ICF tool for categorizing behavioural and physiological data in systematic reviews of digital biomarker-based studies.

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.

II.2 Methods

This section is based on a research paper published at the JMIR mHealth and uHealth journal (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.

II.2.1 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

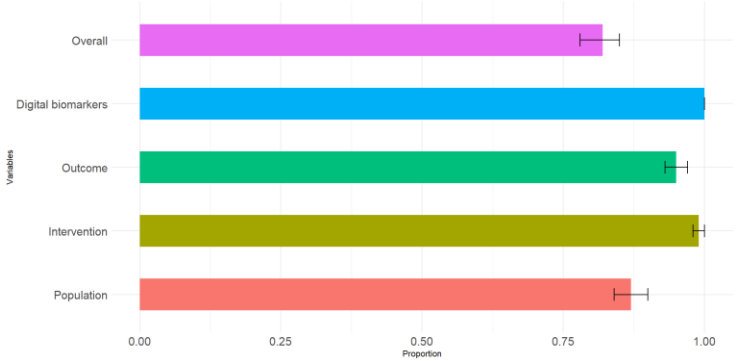
E: All DBMs can be categorized using the ICF tool, otherwise 0 (No).

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.

II.3 Results

273 single studies were considered for the final analysis.

Based on the data presented in the subsequent graph, it was concluded the upper limits for intervention, outcome, and digital biomarker surpass 0.95.



Consequently, we infer that ICHI can be employed to classify interventions, and ICF tool can be used to categorize outcomes and behavioural/physiological data (digital biomarkers) in studies focused on digital biomarkers. However, it is noteworthy that these tools are not suitable for populations, as the upper limit falls below 0.95. Furthermore, upon considering all the study components, where the upper limit is 0.85, we refute the hypothesis (H1) suggesting the appropriateness of WHO tools for categorizing digital biomarker studies. Furthermore, as a result of excluding 35 studies involving healthy and non-clinical populations, the analysis reveals an upper limit of 0.96. Accordingly, we establish that WHO tools are applicable for categorizing digital biomarker studies focusing on patients with specific diseases.

II.4 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.

III. COMPARING THE STATISTICAL POWER OF DIRECT AND INDIRECT DBMs

III.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). 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 regulatory 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.

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. We assumed that direct digital biomarker studies have higher power than that of indirect digital biomarkers. 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.

III.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.

III.2.1 Power calculation

For continuous outcomes the 'pwr' package in the R programming language, and for dichotomous outcomes, the STATA package 'Power' were utilized. All power calculations were performed considering a 95% confidence levels and an alpha value of 0.05.

III.2.2 Statistical analysis

To test the hypothesis, the average of statistical power of the direct and indirect DBMs was calculated and compared. 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. A non-parametric Mann Whitney U test with 95% confidence intervals was used to test the hypothesis.

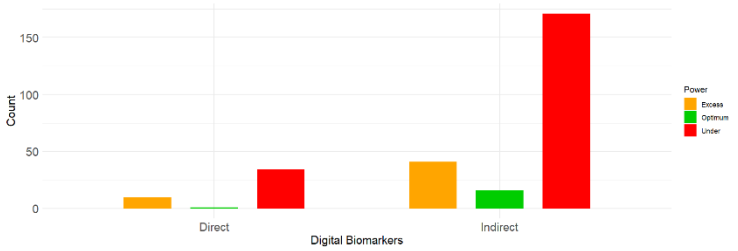
III.3 Results

273 studies were identified. The statistical power of each study was calculated and reported as a measure between 0 and 1.

The analysis for testing the normal distribution showed non-normal distribution of power in both groups direct and indirect digital biomarkers (p-values <0.05). Consequently, the results of Wilcoxon rank-sum (Mann-Whitney) test revealed a non-significant difference between direct and indirect digital biomarkers in terms of the statistical power (p-value>0.05).

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 the following figure, 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.



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

III.4 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.

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

IV.1 Introduction

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, 2010; 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. 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.

IV.2 Methods

The most significant digital biomarker-based systematic reviews (implantable cardiac defibrillators in heart failure patients for mortality and wearable activity trackers to change weight in all populations) and their non-digital biomarker or pharmaceutical pairs were identified through separate searches in the PubMed electronic database.

IV.2.1 Statistical analysis

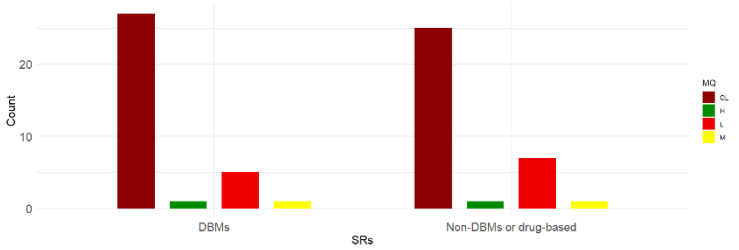
The methodological quality of the included systematic reviews was assessed using the AMSTAR-2 tool (Shea *et al.*, 2017). Consequently, Shapiro Wilk test with a 95% confidence intervals was used to ascertain the normality of the methodological quality between the two groups. After checking the normal distribution of the data between the two groups, a two sample t-test was used to test the hypothesis.

IV.3 Results

20 implantable cardiac defibrillator-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.

Due to the normal distribution of the methodological quality in both groups (p-values >0.05), a t-test revealed a non-significant difference in the methodological quality between the two groups (p-value>0.05). 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.

Subsequent assessments also revealed that the majority of included systematic reviews in both groups exhibited critically low levels of methodological quality, as depicted in the following figure.



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

DBMs: Digital biomarkers, SRs: Systematic reviews, MQ: Methodological quality, H: High methodological quality, M: Moderate methodological quality, L: Low methodological quality, CL: Critically low methodological quality.

IV.4 Conclusion

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.

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.

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

V.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, 2018). 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. We hypothesized that the number of low or very low quality of evidence of digital biomarker-based meta-analyses is significantly higher than that on non-digital biomarker or pharmaceuticals. 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.

V.2 Methods

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

V.2.1 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.

V.2.2 Statistical analysis

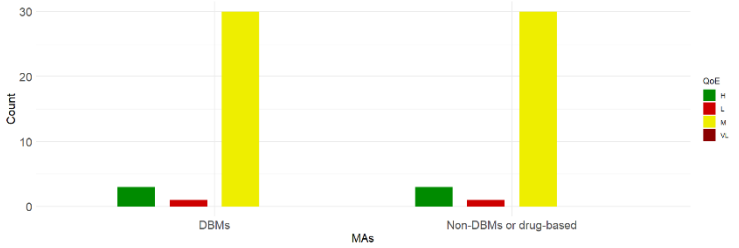
Meta-analyses were divided into two groups: digital biomarkers and non-digital biomarkers or pharmaceuticals, whose quality of evidence was assessed using the above strategies. If the quality of evidence of a meta-analysis was low or very low, we selected "yes". Otherwise (if the quality of evidence 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.

V.3 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.

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. 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 as shown in the following graph.



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.

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.

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). The following Table shows the descriptive statistics of the variables in both groups' digital biomarkers and non-digital biomarkers or drug-based meta-analyses.

The distribution of effect sizes' magnitudes within both groups was also assessed as a supplementary analysis. 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.

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

V.4 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.

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

VI.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.

VI.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 and separate searches were conducted to find their matched non-digital biomarkers or pharmaceutical-based meta-analyses.

IV.2.1 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.

VI.2.2 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.

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 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.

VI.3 Results

20 meta-analyses from 13 SRs which includes at least 10 studies were deemed eligible for the final examination. 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.

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

The outcomes of Mann-Whitney U test failed to reveal any significant disparities between the two groups in terms of either the number of missing studies or the relative changes in effect sizes (p-value >0.05). Therefore, we conclude that no significant distinctions exist between the two groups concerning the number of missing studies and alterations in effect sizes.

VI.4 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 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.

VII. Conclusions and practical implications

VII.1 Classification of DBM studies using WHO tools

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. Our observation on the coverage of populations by WHO systems, revealing an emphasis on

populations with illness while offering limited coverage for healthy populations.

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 (Khan *et al.,* 2017). 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. 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.

VII.2 Statistical power of direct – indirect biomarkers

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). The Medical Device

Regulation (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. 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.

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.

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

Irrespective of variations in industry structure, technology, and regulatory standards, 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. 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.

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.

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.

VII.4 Conclusion

The Medical Device Regulation (MDR) has significantly improved European medical device regulatory standards, addressing the above concerns and improving clinical evidence. Despite MDR implementation delays, digital health technology evidence requirements are rising. Companies that achieve these higher clinical requirements will survive and obtain access to large interconnected markets, while those that fail may lose their market authorisation. Thus, medical technology enterprises may gain a competitive edge by strategically planning and executing extensive clinical investigations

to provide high-quality clinical data. Developing these essential skills needs immediate attention and effort. Digital health investors should actively monitor industry players' evidence quality and clinical trial competence, since these characteristics may significantly increase company risk.

VIII. MY OWN PUBLICATIONS

VIII.1 Journal articles

VIII.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.

VIII.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.

VIII.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.

VIII.2 Chapter in book

VIII.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.

IX. References

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.

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

Bastian, H., Glasziou, P. and Chalmers, I. (2010) ‘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.

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.

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

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.

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

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.

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.

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., 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.

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

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.

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.

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).

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.

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.

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.

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).

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

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).

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.

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

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

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.

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.

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.

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).

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.

Shea, B. J. *et al.* (2017) '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.

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.

Szajewska, H. (2018) '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.

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.

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.

WHO (2017) *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 (2020) *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).

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).

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.