Quality of evidence assessment of direct and indirect digital biomarkers

Hossein Motahari-Nezhad[1], Zsombor Zrubka[2]

[1]Óbuda University, Budapest, Hungary, Doctoral School of Business and Management, Corvinus University of Budapest, Budapest, Hungary motahari.hossein@uni-obuda.hu

[2]Health Economics Research Center, University Research and Innovation Center, Óbuda University, Budapest, Hungary, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary zrubka.zsombor@uni-obuda.hu

Abstract: Wearable, portable, implanted, or ingestible digital devices are used to gather objective, quantifiable, physiological, and behavioral parameters known as digital biomarkers. Meta-analyses of digital biomarkers have increased in recent years, causing the need for evaluating evidence quality. This study evaluated the quality of direct and indirect digital biomarker evidence. Using the definition of digital biomarkers, a search of PubMed and the Cochrane Library was undertaken for articles published in 2019 and 2020. English meta-analyses comparing the clinical effects of therapies using digital biomarkers to those using non-digital biomarker-based interventions were considered. The GRADE technique was used to evaluate the quality of the evidence synthesis in the meta-analyses. The final analysis comprised 26 articles with 95 reported outcomes (26/95, 27.4% direct and 69/95, 72.6% indirect digital biomarkers). There was moderate quality evidence for most indirect digital biomarkers (64/69, 92.8%) and the quality of evidence was low for five outcomes (7.2%). Direct digital biomarkers generated 23.1% (6/26) high-quality evidence and 76.9% (20/26) moderate-quality evidence. Most direct and indirect digital biomarkers provided moderate quality evidence, however, direct biomarkers, because they influence physiological indicators, generate higher evidence.

Keywords: Digital biomarkers; wearable; implantable; quality of evidence; GRADE

1 Introduction

New sensor-based gadgets and wearables have revolutionized the way clinical data is measured, collected, and stored, which has far-reaching implications for medical judgment (Hutchison, 2019). In the medical world, digital biomarkers are a relatively recent notion. They are objective, quantitative, physiological and

behavioral data gathered by portable, wearable, implantable, or digestible digital devices (Babrak et al., 2019). By remotely and continuously measuring reliable clinical data and facilitating continuous monitoring and evaluation, digital biomarkers can increase the diagnostic and therapeutic precision of the current healthcare system (Lipsmeier et al., 2018; Robb et al., 2016). Utilization of digital biomarkers outside of a clinical setting is feasible thanks to the availability of ingestible, implantable, and wearable devices and sensors that can collect data (Byrom et al., 2018).

Consistent with this concept, digital devices collect two different types of data: physiological data and behavioral data, commonly known as direct and indirect interventions. Direct interventions have a direct effect on the physiological characteristics of a patient group, whereas indirect interventions only measure and display the behavioral characteristics of a population (Meerwijk et al., 2016). For instance, the complex electronic devices known as implantable cardioverter defibrillators (ICDs) are used to detect and treat cardiac arrhythmias (Mahdi Abid et al., 2022). The ability of these devices to provide demand-controlled pacing makes it conceivable that they can be used as both emergency defibrillators and backup pacemakers. They continuously monitor the patients' heart rhythm. When the heart rhythm gets out of balance, the pulse generator sends out some shocks to regulate the balance of the heartbeat. In fact, they directly influence and change the physiological data of the patients (Ammannaya, 2020), while physical activity trackers only measure and monitor people's behavioral data; they do not directly influence or change users' physiological characteristics (Shin et al., 2019).

It is getting more challenging for doctors to review and assess all available data in a medical sector because of the steady stream of new medical research being published. In evidence-based medicine, the meta-analysis is considered being the "gold standard" since it provides a more accurate and less biased evaluation of a clinical condition (Lee, 2018). Because of its superiority over subjective judgment as a standard approach for summarizing the findings of several studies, meta-analysis has grown in prominence as a research tool in recent years. While meta-analyses offer significant advantages for scientific research in medicine, they also have some disadvantages. The validity of a meta-analysis can be questioned for several reasons(Lee, 2018). To evaluate the quality of meta-analysis findings, there is a reliable instrument called GRADE (quality of evidence). Thus, it is crucial that all measures be taken to improve the quality of evidence, for which high-quality evidence is necessary. GRADE Quality of evidence (QOE) is defined by the Grading of Recommendations Assessment, Development, and Evaluation Working Group as the level of assurance that the stated impact estimates are reliable and sufficiently competent to warrant the suggestion in question (Guyatt, Oxman, Kunz, et al., 2008). Recently, there has been a rise in the number of published meta-analyses and systematic reviews of digital biomarkers. However, direct and indirect digital biomarkers must be compared to determine which produces the higher evidence. Therefore, this research aims to evaluate the quality of evidence produced

by direct and indirect digital biomarkers to determine which form of digital biomarkers create the highest quality evidence.

2 Methods

Included were meta-analyses of systematic reviews that compared the clinical efficacy of digital biomarker-based interventions to that of non-digital biomarker-based alternatives. In particular, we looked at meta-analyses that aggregated data from DBM research in human populations across all diseases, ages, and sexes. Consideration was given to all initiatives that use DBMs for diagnosing patients, monitoring results, or affecting a therapeutic intervention. As long as the control group didn't use DBMs for the aforementioned goals, the choice of comparator was unrestricted. This review particularly omitted systematic studies that assessed the measurement characteristics or other technical or usage features of DBMs that did not result in a change in the health status of the participants. Full-text publications published in English between January 1, 2019, and December 31, 2020, in peer-reviewed journals were examined. Between 2019 and 2020, we conducted a targeted literature search in PubMed and the Cochrane Library. Also, we searched through the citations of related systematic reviews. The literature search included the terms "digital biomarker," "wearable," "implantable," "portable," and "digestible" in relation to digital biomarkers (Hossein Motahari-Nezhad et al., 2021). After deleting duplicates, the article titles/abstracts and full-texts were evaluated based on the inclusion criteria. After identifying the final qualifying meta-analyses, direct and indirect digital biomarkers were distinguished, as previously showed. Using the GRADE method, the strength of the evidence was evaluated for each outcome (Guyatt, Oxman, Vist, et al., 2008; Schünemann et al., 2013). According to the number of downgrades in each criterion, this tool rates the quality of the evidence as high, moderate, low, or very low, based on five separate factors, including risk of bias, inconsistency, imprecision, publication bias, and indirectness.

If seventy-five percent or more of the included studies matched the criteria for low risk of bias, no downgrading was given to the outcome. If less than 75% of the listed studies had a low risk of bias or if the risk of bias was not specified, a downgrade was given (Pollock et al., 2016). According to the stated heterogeneity for each outcome, inconsistency was evaluated. If the I^2 statistic was less or equal to 75%, there was no downgrade. If the I^2 value exceeded 75% or was not reported, a downgrading was applied. If just one research was considered for the outcome, there was no downgrading (Pollock et al., 2016). The imprecision was determined by analyzing the sample size (Zhang et al., 2019). The weight of the evidence was not downgraded if the combined sample size surpassed 2,000(Schünemann et al., 2013). If the total sample size was less than 200, we degraded the rating by one point. Using STATA 16, we examined the optimum information size (OIS) for sample sizes between 200 and 2000 using power analysis. If the result of the power analysis exceeded the stated number of participants in the outcome, one downgrade was

given; otherwise, no downgrade was implemented. Regarding publication bias, the trim and fill technique was used (Duval & Tweedie, 2000). Differences between the included studies and the meta-analysis's question were considered when determining indirectness for each outcome.

3 Results

There were 389 studies found while searching PubMed and the Cochrane Library. 375 studies met the criteria for title/abstract screening after removing duplicates. Full-text screening included 199 studies. In all, 175 full-text articles were disqualified because they did not meet the criterion. Two more systematic reviews that satisfied the inclusion criteria were discovered by checking the reference lists of the eligible papers. Therefore, 26 systematic reviews were accounted for in the results. There were a total of 26 reviews included for this study, including 95 outcomes in the reviews, with five of them being considered to be of low quality (5%), eighty-four being considered being of moderate quality (89%), and six being considered to be of high quality (6%). Most outcomes had some risk of bias (89/95, 94%), followed by inconsistency (27/95, 28%), and imprecision (27/95, 28%), when compared to the GRADE criteria. Only two out of ninety-five outcomes (2%) showed evidence of publication bias. However, the outcomes did demonstrate any indirectness. Furthermore, 70 outcomes (74% of total) did not have publication bias checked because there were not enough studies meeting the inclusion criteria. There were additional three outcomes (3%) that could not be checked for inconsistency since just one research met the inclusion criteria.

It was found that most outcomes (69/95, 73%) were identified as indirect digital biomarkers, whereas only 26/95 (27%), were categorized as direct digital biomarkers. The majority of indirect digital biomarkers were classified as having moderate evidence quality (64/69, 93%), whereas the number of indirect digital biomarkers with low evidence quality was found to be 5 (7%). No indirect digital biomarkers with high-quality evidence were identified. Regarding direct digital biomarkers, most of the evidence was of moderate quality (20/26, 77%), whereas no direct digital biomarkers were classified as having evidence of low quality. In addition, 23% (6/26) of direct digital biomarkers were classified as having high-quality evidence.

4 Discussion

In this research, direct and indirect digital biomarkers' evidence quality was evaluated using the GRADE instrument. A thorough search was undertaken in PubMed and the Cochrane Library electronic databases to identify systematic

reviews, including meta-analyses evaluating the influence of digital biomarkers on health outcomes relative to non-digital biomarker comparators. Finally, 26 systematic reviews with 95 outcomes were discovered and included. No indirect digital biomarkers with high evidence quality were discovered, however, 23% of direct digital biomarkers were evaluated as having high evidence quality. The results suggested that the influence of an implantable cardiac defibrillator as a direct digital biomarker on all-cause mortality was validated by high-quality data for ICDs implanted both before and after CF-LVAD placement. In addition, based on the analysis, we have a very high level of confidence on the influence that the implantable cardiac defibrillator as a direct digital biomarker has on the likelihood of receiving a transplant, the rate at which atrial arrhythmias are detected, and the occurrence of strokes. The influence of indirect digital biomarkers, such as wearable activity monitors, on steps in chronic pulmonary disease, as well as steps in obese and sedentary older individuals, was assessed to have low-quality evidence. According to the findings, direct digital biomarkers provide evidence of a higher quality than indirect digital biomarkers. Physical activity trackers, the most common form of indirect digital biomarkers (H Motahari-Nezhad et al., 2022), are extensively used by the general population (Macridis et al., 2018), particularly to increase physical activity levels (Bentley et al., 2020). Over the course of the last decade, there has been a substantial growth in the market for wearable activity trackers all over the globe. Between the years 2014 and 2020, the number of shipments of wearable activity trackers around the globe skyrocketed (Statista, 2022). Despite their initial promise, there is substantial skepticism in the scientific, medical, and public communities over the efficacy of wearable activity monitors (Lieberman, 2015). Indirect digital biomarkers have a large market and user base, and as a result, they must deliver more reliable level of evidence; otherwise, their usefulness should be questioned (Ferguson et al., 2022). To ensure the success of digital health technologies in improving the health of its users, it is essential that managers, engineers, and healthcare experts and physicians work together to develop these tools (Gondi et al., 2022).

Conclusion

In general, the evidence provided by digital biomarkers is of a moderate quality. When comparing the evidence provided by direct and indirect digital biomarkers, the findings suggest that direct digital biomarkers give a greater evidence.

Acknowledgements

I wish to acknowledge the help provided by Professor László Gulácsi (Health Economics Research Center, University Research and Innovation Center, Obuda University, Budapest, Hungary; Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary), and professor Márta Péntek (Health Economics Research Center, University Research and Innovation Center, Obuda University, Budapest, Hungary). This project was implemented under the Thematic Excellence Programme (TKP), supported by the National Research, Development, and Innovation Office. Project ID: TKP2021-NKTA-36.

References

[1] Ammannaya, G. K. K. (2020). Implantable cardioverter defibrillators – the past, present and future. *Archives of Medical Science – Atherosclerotic Diseases*, *5*(1), 163–170. https://doi.org/10.5114/amsad.2020.97103

[2] Babrak, L. M., Menetski, J., Rebhan, M., Nisato, G., Zinggeler, M., Brasier, N., Baerenfaller, K., Brenzikofer, T., Baltzer, L., Vogler, C., Gschwind, L., Schneider, C., Streiff, F., Groenen, P. M. A., & Miho, E. (2019). Traditional and Digital Biomarkers: Two Worlds Apart? *Digital Biomarkers*, *3*(2), 92–102. https://doi.org/10.1159/000502000

[3] Bentley, C. L., Powell, L., Potter, S., Parker, J., Mountain, G. A., Bartlett, Y. K., Farwer, J., O'Connor, C., Burns, J., Cresswell, R. L., Dunn, H. D., & Hawley, M. S. (2020). The Use of a Smartphone App and an Activity Tracker to Promote Physical Activity in the Management of Chronic Obstructive Pulmonary Disease: Randomized Controlled Feasibility Study. *JMIR MHealth and UHealth*, *8*(6), e16203. https://doi.org/10.2196/16203

[4] Byrom, B., Watson, C., Doll, H., Coons, S. J., Eremenco, S., Ballinger, R., Mc Carthy, M., Crescioni, M., O'Donohoe, P., & Howry, C. (2018). Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. *Value in Health*, *21*(6), 631–639. https://doi.org/10.1016/j.jval.2017.09.012

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

[6] Ferguson, T., Olds, T., Curtis, R., Blake, H., Crozier, A. J., Dankiw, K., Dumuid, D., Kasai, D., O'Connor, E., Virgara, R., & Maher, C. (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), e615–e626. https://doi.org/10.1016/S2589-7500(22)00111-X

[7] Gondi, S., Powers, B. W., & Shrank, W. H. (2022). Evidence and Efficacy in the Era of Digital Care. *Journal of General Internal Medicine*, *37*(10), 2559–2561. https://doi.org/10.1007/s11606-022-07445-0

[8] Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is "quality of evidence" and why is it important to clinicians? *BMJ*, *336*(7651), 995–998. https://doi.org/10.1136/bmj.39490.551019.BE

[9] Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

[10] Hutchison, D. (2019). *High-Performance Modelling and Simulation for*

Big Data Applications (J. Kołodziej & H. González-Vélez (eds.); Vol. 11400). Springer International Publishing. https://doi.org/10.1007/978-3-030-16272-6

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

[12] Lieberman, B. (2015). *The 4 Biggest Problems With Your Fitness Tracker, According To Scientists*. https://www.self.com/story/fitness-tracker-accuracy-wearables

[13] Lipsmeier, F., Taylor, K. I., Kilchenmann, T., Wolf, D., Scotland, A., Schjodt-Eriksen, J., Cheng, W.-Y., Fernandez-Garcia, I., Siebourg-Polster, J., Jin, L., Soto, J., Verselis, L., Boess, F., Koller, M., Grundman, M., Monsch, A. U., Postuma, R. B., Ghosh, A., Kremer, T., … Lindemann, M. (2018). Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Movement Disorders*, *33*(8), 1287–1297. https://doi.org/10.1002/mds.27376

[14] Macridis, S., Johnston, N., Johnson, S., & Vallance, J. K. (2018). Consumer physical activity tracking device ownership and use among a population-based sample of adults. *PLOS ONE*, *13*(1), e0189298. https://doi.org/10.1371/journal.pone.0189298

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

[16] Meerwijk, E. L., Parekh, A., Oquendo, M. A., Allen, I. E., Franck, L. S., & Lee, K. A. (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), 544–554. https://doi.org/10.1016/S2215-0366(16)00064-X

[17] Motahari-Nezhad, H, Fgaier, M., Péntek, M., Gulácsi, L., & Zrubka, Z. (2022). MT14 Populations, Interventions, and Outcomes in Digital Biomarker-Based Interventions' Systematic Reviews: A Scoping Review. *Value in Health*, *25*(7), S534. https://doi.org/10.1016/j.jval.2022.04.1293

[18] Motahari-Nezhad, Hossein, Péntek, M., Gulácsi, L., & Zrubka, Z. (2021). Outcomes of Digital Biomarker–Based Interventions: Protocol for a Systematic Review of Systematic Reviews. *JMIR Research Protocols*, *10*(11), e28204. https://doi.org/10.2196/28204

[19] Pollock, A., Farmer, S. E., Brady, M. C., Langhorne, P., Mead, G. E., Mehrholz, J., Van Wijck, F., & Wiffen, P. J. (2016). An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. *Journal of Clinical Epidemiology*, *70*, 106–110. https://doi.org/10.1016/j.jclinepi.2015.08.013

[20] Robb, M. A., McInnes, P. M., & Califf, R. M. (2016). Biomarkers and

Surrogate Endpoints. *JAMA*, *315*(11), 1107. https://doi.org/10.1001/jama.2016.2240

[21] Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). *Introduction to GRADE Handbook Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach*. https://gdt.gradepro.org/app/handbook/handbook.html#h.9rdbelsnu4iy

[22] Shin, G., Jarrahi, M. H., Fei, Y., Karami, A., Gafinowitz, N., Byun, A., & Lu, X. (2019). Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. *Journal of Biomedical Informatics*, *93*, 103153. https://doi.org/10.1016/j.jbi.2019.103153

[23] Statista. (2022). *Forecast unit shipments of wrist-worn wearables worldwide from 2019 to 2024(in million units)*. https://www.statista.com/statistics/296565/wearables-worldwide-shipments/

[24] Zhang, Y., Coello, P. A., Guyatt, G. H., Yepes-Nuñez, J. J., Akl, E. A., Hazlewood, G., Pardo-Hernandez, H., Etxeandia-Ikobaltzeta, I., Qaseem, A., Williams, J. W., Tugwell, P., Flottorp, S., Chang, Y., Zhang, Y., Mustafa, R. A., Rojas, M. X., Xie, F., & Schünemann, H. J. (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*, 83–93. https://doi.org/10.1016/j.jclinepi.2018.05.011