Development of machine learning: derived mhealth composite biomarkers for trial@home clinical trials

Zhuparris, A.

PART I

INTRODUCTION

CHAPTER 1

Introduction

Development of novel biomarkers

Clinical biomarkers serve a critical role in diagnosing diseases, monitoring disease progression, measuring drug effects, and predicting treatment outcomes.[1] As our understanding of biology and diseases continue to evolve, there is a growing demand for the development of novel biomarkers that offer more precise, in-depth, and timely understanding of the disease and provide early detection and quantification of drug effects. To meet this need, researchers are increasingly turning towards novel technologies that enable the development of innovative biomarkers. This goal is not without hurdles. Challenges such as data collection, standardization, validation, and regulatory considerations need to be carefully addressed. Additionally, the translation of these biomarkers from research setting to clinical practice requires robust evidence of their clinical utility and reliability.

The primary objective of this thesis is to address the development and validation of innovative biomarkers by harnessing the data of mobile health (MHEALTH) devices, such as smartphones, tablets, and wearable devices. These widely available and data-intensive technologies offer an unprecedented opportunity to capture diverse physiological and behavioral data outside the traditional clinical setting. To effectively utilize this wealth of information, Machine Learning (ML) techniques will be employed to transform the unstructured and multifaceted MHEALTH data into meaningful clinical biomarkers. This research aims to address the challenges, important factors, and potential benefits associated with the development and validation of MHEALTH biomarkers.

MHEALTH devices for clinical trials

Clinical trials play a crucial role in assessing the efficacy of new pharmacological treatments and are typically conducted by academic hospitals and Contract Research Organizations (CROS). Conventionally, data for observational and randomized clinical trials is collected during patients' visits to in-patient facilities like hospitals or clinical research units. This approach has several benefits, such as strict control over the study environment and standardized data collection. However, a limitation is that the data collected only represents a snapshot of the patient's health and disease activity, often in an isolated context. As a result, evidence gaps between visits are created, and clinicians' insight into patients' overall health may be limited.

To overcome the limitations of conventional clinical trials, MHEALTH devices like smartphones, wearables, and tablets offer a unique opportunity for continuous and longitudinal data collection from clinical trial participants under free-living conditions.[2–6] Mobile applications (apps) installed on smartphones and tablets can be utilized to actively collect self-reported outcomes from patients through electronic diaries.[7] Simultaneously, apps can passively collect data from various sensors such as accelerometers, cameras, gyroscopes, microphones, and phone logs, providing an additional source of valuable physical and behavioral data.[8–10] Wearables support continuous tracking of physiological responses or physical activity, such as heart rate or steps, enable characterization of intra- and inter-individual variability in disease activity and quantification of drug response.[11–14] This approach of collecting data from multiple sensors acknowledges that a patient's experience of their disease is a consequence of multiple neurobiological processes, and therefore is expressed as a diverse array of symptoms simultaneously.

The use of MHEALTH devices in clinical trials has sharply increased since the global adoption of the smartphone. Between 2012 to 2022, the term 'MHEALTH' was incorporated in 1605 clinical studies posted on clinicaltrials.gov. Only 15 studies used the term between 2000 to 2011.[15] MHEALTH biomarkers have been shown to be effective in monitoring disease activity and estimating symptom severity for a wide range of diseases such as mood disorders,[16–21] neurodegenerative disorders,[22–24] and cardiovascular diseases.[25] The benefits of MHEALTH devices in clinical trials are twofold. First, real-world data collected under free-living conditions, which is data collected outside of controlled clinical trial settings, can be used to

generate novel hypotheses or insights into the most effective treatments. This can help to provide the ecological validity of findings produced by well-controlled clinical trials. Second, the use of MHEALTH devices for clinical trials may also be cost-effective due to the emerging concept of Bring Your Own Device (BYOD).[26,27] By leveraging participants' own devices for data collection, costs are reduced for clinical trials as study specific hardware does not need to be purchased, distributed, or maintained. The burden for participants is also reduced as they can use hardware that they are already familiar with and can have access to in their daily lives.

Despite these advantages, integrating MHEALTH devices into clinical trials presents its own challenges. The most significant issues include ensuring tolerability and usability of the MHEALTH devices by patients and clinicians and developing, validating, and interpreting the biomarkers given the lack of control under free-living conditions.[5] Unlike controlled clinical settings, free-living conditions offer minimal control over the environment in which data is collected. Participants may also engage in various activities and encounter unpredictable situations that can influence data quality and consistency. Factors such as variations in daily routines, social interactions, and environmental exposures can introduce variability and noise into the collected data. The accuracy and reliability of the collected data can be affected by factors such as user engagement, device performance, and data synchronization. Ensuring data quality requires clear patient instructions, participant compliance, and regular monitoring to address any issues that may arise. When collecting data in free-living conditions, there is a greater risk of breaching participants' privacy. The use of MHEALTH devices, such as smartphones and wearable devices, often involves capturing personal information and sensitive data. Safeguarding privacy becomes crucial to ensure participants' trust and compliance. Implementing robust data encryption, secure data storage, and strict privacy policies are essential to mitigate privacy risks. The datasets generated by these devices are often complex, large, and subject to influence by external factors such as differences in devices, lifestyles, weather, and location. ML provides a potential solution for processing these large and heterogeneous datasets into biomarkers that can aid the understanding and prediction of complex clinical outcomes.

ML and traditional statistical learning methods both play important roles in the analysis and interpretation of clinical trial data. While both share a common objective of extracting meaningful insights and informing decision-making, they have distinct approaches and applications.[28] Traditional statistical learning methods typically focus on hypothesis testing, parameter estimation, and model interpretability and inference and therefore are classically used to test the significance of individual covariates or predictors, estimating effect sizes, and calculating sample sizes.[29] As traditional statistical learning methods are typically designed to answer specific research questions or test predefined hypotheses, their primary focus is on estimating the effects of individual covariates or predictors rather than generating accurate predictions for new, unseen data. These methods may lack the ability to generalize well to different populations, settings, or contexts, as they are often tailored to the specific characteristics of the analyzed dataset. With time-honored techniques such as ANOVA, t-tests, linear and logistic regression, and survival analysis deeply rooted in the field of clinical trials, the continued utilization of traditional statistical learning remains pivotal in advancing medical research and improving patient outcomes.[29,30] However, their limitations can hinder their effectiveness in analyzing complex and diverse clinical trial data, where flexibility and adaptability may be required.

Conversely, ML is primarily focused on developing data-driven statistical models that are both generalizable and predictive in nature.[28,31,32] As a result, ML is often considered more 'data-hungry' compared to statistical learning due to its reliance on large and diverse datasets. Generalizability is a desirable characteristic of biomarkers as it indicates their ability to perform well in diverse scenarios. Generalizable and predictive biomarkers derived from ML techniques can be applied across different patient populations, settings, and clinical trial protocols. A key step in the ML pipeline is the use of cross-validation. By employing cross-validation, clinicians can obtain a reliable estimate of how well the ML model is likely to perform

on unseen data sourced from a similar population or setting. This assessment of predictive accuracy is crucial in determining whether the developed model can generalize its findings beyond the specific dataset used for training. This versatility allows for the broader utilization of biomarkers in various healthcare contexts, increasing their potential impact and value.

A ML model has the potential to build a representative composite biomarker by integrating and capturing complex relationships among different features, which would lead to a more comprehensive and informative representation of the underlying biological or pharmacological processes. However, while the complexity of the biomarker can increase its predictive accuracy, it may limit its interpretability. ML offers a wide range of model types, such as decision trees, neural networks, ensemble methods, transfer learning, and unsupervised learning methods that can be adapted to different types of data and objectives, allowing for more flexible and adaptable modelling approaches.[28,33] Many ML algorithms, particularly deep learning models, can automatically learn and extract features directly from the data, eliminating the need for manual feature engineering. The automation of the identification of relevant features and patterns in the data, reduces the need for manual feature selection and engineering. This can streamline the biomarker development process and improve the efficiency of clinical trial analyses. In addition, unsupervised learning algorithms, which can identify patterns in data without being explicitly told what to look for, can be useful for exploratory data analysis or for discovering hidden patterns or subgroups within data that may not be immediately apparent.[34] In conclusion, ML's data-driven approach, flexibility in model selection, automated feature extraction, and ability to identify hidden patterns offer significant advantages over traditional statistical learning methods in the development of biomarkers for clinical trials. Its reliance on large and diverse datasets may make it more data-hungry, but this enables the creation of generalizable and predictive models. By streamlining the biomarker development process and improving the efficiency of clinical trial analyses, ML has the potential to greatly impact clinical research and contribute to improved patient outcomes.

Clinical validation of composite MHEALTH biomarkers

Composite MHEALTH biomarkers can offer several benefits to both clinicians and patients. By consolidating multiple clinical features into a single composite digital biomarker, this biomarker can be used to predict clinical outcomes, serving as a complement rather than a replacement for multiple clinical endpoints. The resulting composite biomarkers have the potential for inference and prediction, contributing to the discovery of generalizable and robust evidence to guide clinical studies. This thesis proposes that there are three beneficial applications for composite biomarkers. Firstly, composite biomarkers may be more sensitive to subtle changes or treatment effects that may not be evident when assessing individual biomarkers independently. Secondly, by combining multiple biomarkers, this can help mitigate the measurement variability that are inherent in an individual biomarker. The aggregated biomarker can provide a more stable representation of the underlying phenomenon. Lastly, a composite biomarker may provide a more holistic evaluation of disease activity. A composite biomarker provides a more comprehensive and multi-faceted assessment, and therefore may capture a broader spectrum of treatment effects. However, to determine if these composite digital biomarkers have utility in clinical research, they must be clinically validated.[35] The following section addresses the validation criteria considered to evaluate if a biomarker is suitable for clinical adoption.

Validation of novel composite biomarkers before incorporating them into clinical trials is crucial. To validate these biomarkers, Kruizinga et al. have proposed five criteria, which we have adopted along with an optional criterion of Interpretability and Explainability.[35] The first criterion, *Classifying Patients and Healthy Controls*, focuses on accurately distinguishing between patients and healthy individuals to identify disease-specific biomarkers. The second criterion, *Correlation with Gold Standard or Disease Metrics*, involves establishing the validity of the biomarker and its ability to accurately reflect disease activity by correlating it with the

gold standards. The third criterion, *Detecting Changes in Disease Activity or Treatment Effects,* refers to detecting changes in disease activity over time, which is crucial for monitoring disease progression or response to treatment. The fourth criterion, *Tolerability and Usability*, is particularly important for MHEALTH devices that may be worn continuously or for extended periods. The device should not cause discomfort or irritation and should be easy to use. If tolerability and usability of the device are poor, the missing or poor-quality data collected will negatively impact the development of the biomarker. The fifth criterion, *Repeatability and Variability*, refers to the device producing consistent measurements under different conditions and over multiple time points. Finally, the optional criterion, *Interpretability and Explainability*, refers to the ability of the composite biomarker to provide clear and understandable explanations for its predictions. This is important for building trust in the biomarker and its ability to inform clinical decision-making.

Research objectives and structure of this thesis

The overall research question of this thesis is *How can MHEALTH devices and ML algorithms be used to develop composite biomarkers for clinical applications*? To address this question, we have outlined a series of research questions that will explore different aspects of the development and clinical validation of these biomarkers. These research questions will be addressed in their respective chapters, culminating in a discussion of the general findings and recommendations for future research in this field.

Parts 2 to 4 will use clinical trial data collected using Centre for Human Drug Research (CHDR)'s Trial@Home platform. The Trial@Home platform aims to investigate alternative approaches for collecting clinical trial data in non-traditional clinical settings. Serving as a comprehensive solution, Trial@Home offers end-to-end services, encompassing trial design, execution, and data analytics. By integrating smartphones, tablets, and wearables (such as smartwatches, smart scales, and sleep mats) into clinical trials, participants can experience reduced visit frequency while enabling more convenient and representative data collection. This innovative approach captures participants' real-world experiences in their daily lives, providing valuable insights under free-living conditions. Through the use of ML, the collected data is transformed into novel and validated digital biomarkers. The following chapters provide more insight into the type of data collected during these trials, and how the data was transformed into validated biomarkers for clinical applications.

Part 1 (Introduction) asks *What is the motivation behind creating composite MHEALTH biomarkers for clinical applications and how are they currently being developed?* This part addresses the challenges and limitations of using MHEALTH devices and ML for developing and validating composite biomarkers in clinical trials. **Chapter 1** provides a brief overview of concept, reasoning, and importance of using ML in clinical trials that use MHEALTH devices. **Chapter 2** offers a literature review of existing published studies that have used similar techniques to derive composite biomarkers. Given the rise and breadth of ML applications in clinical trials, we sought to identify both the generic and best practices of developing these ML applications. However, given the lack of consistent reporting in these studies, the literature review does not provide a complete or detailed overview. On the contrary, the literature review presents a set of recommended reporting practices aimed at enhancing the transparency and reproducibility of the methods utilized.

Part 2 (Classification of Diagnosis) asks *How can MHEALTH devices and ML be utilized to create composite biomarkers for the classification of diagnoses?* This part addresses how different types of MHEALTH devices compare in terms of their usability, tolerability, and data quality for developing composite biomarkers. Further, it examines the methods required for developing accurate and clinically relevant biomarkers for the classification of disease diagnoses using MHEALTH data and ML. **Chapter 3** use the Trial@Home platform to classify the remotely monitored behavioural activity of Facioscapulohumeral Muscular Dystrophy (FSHD) patients respectively from Healthy Controls. To assess the feasibility of piloting a Trial@Home study, these publications also report the data completion rate and patient experience of the Trial@Home app to reflect the tolerability and usability of the devices.

Part 3 (Estimation of Symptom Severity) asks *How can MHEALTH devices and ML be utilized to create composite biomarkers for the estimation of symptom severity?* This part investigates the effectiveness of the developed composite biomarkers in estimating the severity of disease symptoms in patients compared to traditional methods. **Chapter 4** and **5** use regression algorithms and the Trial@Home platform to estimate the symptom severity of the FSHD and Major Depressive Disorder (MDD) patients. In addition to estimating the symptom severity, we evaluated how varying time windows used to train the models can affect the repeatability and variability of their predicted outcomes. **Chapter 6** and **7** focus on developing ML models that can automatically quantify the number of coughs and cries using a smartphone microphone respectively. While these activities cannot be used as diagnostic tools themselves, they serve as relevant and informative proxies for disease activity.

Part 4 (Detection of Treatment Effects) asks *Can the use of MHEALTH devices and ML algorithms enable the detection of treatment effects in clinical trials and provide* insights *into the efficacy of pharmacological treatments*? To address this question, **Chapter 8** explore if a composite tapping biomarker can detect treatment effects and to estimate symptom severity among Parkinson's Disease patients respectively. The underlying motivation for this investigation lies in examining whether the same tapping biomarker can serve the dual purpose of monitoring both treatment effects and symptom severity in alignment with the gold standard, thus unveiling new possibilities for comprehensive biomarker applications.

Chapter 9, the discussion, reflects on the methodologies and analyses in **Parts 2 to 4** and addresses the motivations, factors, and limitations that contribute to the development and adoption of MHEALTH composite biomarkers for the purposes of diagnosis classification, symptom severity estimation, and treatment effects detection. Given the potential impacts of MHEALTH biomarkers, the discussion reflects on the practical and ethical implications of MHEALTH biomarkers for clinicians, other Central Nervous System (CNS) disorders, and future clinical trials.

Condensed structure of the thesis

Given the criteria for evaluating the clinical validity of candidate composite biomarkers, this thesis consists of 5 parts. Part 1 provides the theoretical and historical framework for the development of these biomarkers. Part 2, 3, and 4 focus on clinical trials that use ML to classify a clinical diagnosis, to estimate symptom severity, and to detect treatment effects respectively. In each of these sections, we provide a detailed account of our approach to the proposed clinical validation. **Chapter 9** discusses the general findings of this thesis and addresses general recommendations for developing future biomarkers that use MHEALTH devices and ML.

REFERENCES

1 A. F. Cohen, J. Burggraaf, J. M. A. van Gerven, M. Moerland, and G. J. Groeneveld, 'The Use of Biomarkers in Human Pharmacology (Phase I) Studies,' *Annu Rev Pharmacol Toxicol*, vol. 55, no. 1, pp. 55–74, Jan. 2015, doi: 10.1146/annurev-pharmtox-011613-135918.

2 D. G. Stuijt, V. Exadaktylos, A. D. Bins, J. J. Bosch, and M. G. H. van Oijen, 'Potential Role of Smartphone-Based Passive Sensing in Remote Monitoring of Patients With Cancer,' *JCO Clin Cancer Inform*, no. 6, Sep. 2022, doi: 10.1200/CCI.22.00079.

3 M. M. E. Vogel, S. E. Combs, and K. A. Kessel, 'MHEALTH and Application Technology Supporting Clinical Trials: Today's Limitations and Future Perspective of smartRCTs,' *Front Oncol*, vol. 7, Mar. 2017, doi: 10.3389/fonc.2017.00037.

4 S. P. Rowland, J. E. Fitzgerald, T. Holme, J. Powell, and A. McGregor, 'What is the clinical value of MHEALTH for patients?,' *NPJ Digit Med*, vol. 3, no. 1, p. 4, Jan. 2020, doi: 10.1038/s41746-019-0206-x.

5 A. Kakkar, P. Sarma, and B. Medhi, 'MHEALTH technologies in clinical trials: Opportunities and challenges,' *Indian J Pharmacol*, vol. 50, no. 3, p. 105, 2018, doi: 10.4103/ijp.IJP_391_18.

6 Q. Pham, D. Wiljer, and J. A. Cafazzo, 'Beyond the Randomized Controlled Trial: A Review of Alternatives in MHEALTH Clinical Trial Methods,' *JMIR MHEALTH Uhealth*, vol. 4, no. 3, p. e107, Sep. 2016, doi: 10.2196/MHEALTH.5720.

7 S. J. Coons, S. Eremenco, J. J. Lundy, P. O'Donohoe, H. O'Gorman, and W. Malizia, 'Capturing Patient-Reported Outcome (PRO) Data Electronically: The Past, Present, and Promise of ePRO Measurement in Clinical Trials,' *The Patient - Patient-Centered Outcomes Research*, vol. 8, no. 4, pp. 301–309, Aug. 2015, doi: 10.1007/s40271-014-0090-z.

8 R. Au, H. Lin, and V. B. Kolachalama, 'Tele-Trials, Remote Monitoring, and Trial Technology for Alzheimer's Disease Clinical Trials,' in *Alzheimer's Disease Drug Development*, Cambridge University Press, 2022, pp. 292–300. doi: 10.1017/9781108975759.026.

9 Z. W. Adams, E. A. McClure, K. M. Gray, C. K. Danielson, F. A. Treiber, and K. J. Ruggiero, 'Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research,' *J Psychiatr Res*, vol. 85, pp. 1–14, Feb. 2017, doi: 10.1016/j.jpsychires.2016.10.019.

10 A. Zhuparris *et al.*, 'Smartphone and Wearable Sensors for the Estimation of Facioscapulohumeral Muscular Dystrophy Disease Severity: Cross-sectional Study,' *JMIR Form Res*, vol. 7, p. e41178, Mar. 2023, doi: 10.2196/41178.

11 G. Cosoli, A. Poli, S. Spinsante, and L. Scalise, 'The importance of physiological data variability in wearable devices for digital health applications,' *ACTA IMEKO*, vol. 11, no. 2, p. 1, May 2022, doi: 10.21014/acta_imeko.v11i2.1135.

12 S. Carreiro *et al.*, 'Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study,' *Journal of Medical Toxicology*, vol. 11, no. 1, pp. 73–79, Mar. 2015, doi: 10.1007/s13181-014-0439-7.

13 A. K. Yetisen, J. L. Martinez-Hurtado, B. Ünal, A. Khademhosseini, and H. Butt, 'Wearables in Medicine,' *Advanced Materials*, vol. 30, no. 33, p. 1706910, Aug. 2018, doi: 10.1002/adma.201706910.

14 S. Bian, B. Zhu, G. Rong, and M. Sawan, 'Towards wearable and implantable continuous drug monitoring: A review,' *J Pharm Anal*, vol. 11, no. 1, pp. 1–14, Feb. 2021, doi: 10.1016/j.jpha.2020.08.001.

15 NIH U.S. National Library of Medicine, 'Information on Clinical Trials and Human Research Studies,' Apr. 01, 2023. https://www.clinicaltrials.gov/ct2/results?cond=&term=MHEALTH&type=&rslt=&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=&state=&city=&dist=&locn=&rsub=&strd_s=01%2F01%2F2012&strd_e=12%2F31%2F2022&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=&rfpd_e=&lupd_s=&lupd_e=&sort= (accessed Apr. 05, 2023).

16 E. Brietzke, E. R. Hawken, M. Idzikowski, J. Pong, S. H. Kennedy, and C. N. Soares, 'Integrating digital phenotyping in clinical characterization of individuals with mood disorders,' *Neurosci Biobehav Rev*, vol. 104, no. July, pp. 223–230, Sep. 2019, doi: 10.1016/j.neubiorev.2019.07.009.

17 A. Grünerbl *et al.*, 'Smartphone-based recognition of states and state changes in bipolar disorder patients,' IEEE *J Biomed Health Inform*, vol. 19, no. 1, pp. 140–148, Jan. 2015, doi: 10.1109/JBHI.2014.2343154.

18 S. Stanislaus *et al.*, 'Smartphone-based activity measurements in patients with newly diagnosed bipolar disorder, unaffected relatives and control individuals,' *Int J Bipolar Disord*, vol. 8, no. 1, p. 32, Dec. 2020, doi: 10.1186/s40345-020-00195-0.

19 M. Faurholt-Jepsen *et al.*, 'Differences in mood instability in patients with bipolar disorder type I and II: a smartphone-based study,' *Int J Bipolar Disord*, vol. 7, no. 1, 2019, doi: 10.1186/s40345-019-0141-4.

20 J.-Y. Liu, K.-K. Xu, G.-L. Zhu, Q.-Q. Zhang, and X.-M. Li, 'Effects of smartphone-based interventions and monitoring on bipolar disorder: A systematic review and meta-analysis,' *World J Psychiatry*, vol. 10, no. 11, pp. 272–285, 2020, doi: 10.5498/wjp.v10.i11.272.

21 M. Faurholt-Jepsen, M. Frost, E. M. Christensen, J. E. Bardram, M. Vinberg, and L. V. Kessing, 'The effect of smartphone-based monitoring on illness activity in bipolar disorder: The MONARCA II randomized controlled single-blinded trial,' *Psychol Med*, vol. 50, no. 5, pp. 838–848, 2020, doi: 10.1017/S0033291719000710.

22 L. C. Kourtis, O. B. Regele, J. M. Wright, and G. B. Jones, 'Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity,' *NPJ Digit Med*, vol. 2, no. 1, pp. 1–9, 2019, doi: 10.1038/s41746-019-0084-2.

23 R. Bhidayasiri and Z. Mari, 'Digital phenotyping in Parkinson's disease: Empowering neurologists for measurement-based care,' *Parkinsonism and Related Disorders*, vol. 80. Elsevier Ltd, pp. 35–40, Nov. 01, 2020. doi: 10.1016/j.parkreldis.2020.08.038.

24 P. Kassavetis *et al.*, 'Developing a Tool for Remote Digital Assessment of Parkinson's Disease,' *Mov Disord Clin Pract*, vol. 3, no. 1, pp. 59–64, Jan. 2016, doi: 10.1002/mdc3.12239.

25 J. X. Teo *et al.*, 'Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging,' *bioRxiv*, 2019, doi: 10.1101/527077.

26 L. Pugliese, O. Crowley, M. Woodriff, V. Lam, J. Sohn, and S. Bradley, 'Feasibility of the 'Bring Your Own Device' Model in Clinical Research: Results from a Randomized Controlled Pilot Study of a Mobile Patient Engagement Tool,' *Cureus*, Mar. 2016, doi: 10.7759/cureus.535.

27 C. Demanuele *et al.*, 'Considerations for Conducting Bring Your Own 'Device' (BYOD) Clinical Studies,' *Digit Biomark*, vol. 6, no. 2, pp. 47–60, Jul. 2022, doi: 10.1159/000525080.

28 I. El Naqa and M. J. Murphy, 'What Is Machine Learning?,' in *Machine Learning in Radiation Oncology*, Springer International Publishing, 2015, pp. 3–11. doi: 10.1007/978-3-319-18305-3_1.

29 R. J. Calin-Jageman, 'Better Inference in Neuroscience: Test Less, Estimate More,' *The Journal of Neuroscience*, vol. 42, no. 45, pp. 8427–8431, Nov. 2022, doi: 10.1523/JNEUROSCI.1133-22.2022.

30 R. Iniesta, D. Stahl, and P. Mcguffin, 'Machine learning, statistical learning and the future of biological research in psychiatry,' 2016, doi: 10.1017/S0033291716001367.

31 T. Hastie, R. Tibshirani, and J. Friedman, 'Statistics The Elements of Statistical Learning,' *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2009.

32 T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York City: Springer Science & Business Media, 2013.

33 D. Bzdok, N. Altman, and M. Krzywinski, 'Statistics versus machine learning,' *Nat Methods*, vol. 15, no. 4, pp. 233–234, Apr. 2018, doi: 10.1038/nmeth.4642.

34 Z. Ghahramani, 'Unsupervised Learning,' 2004, pp. 72–112. doi: 10.1007/978-3-540-28650-9_5.

35 M. D. Kruizinga *et al.*, 'Development of Novel, Value-Based, Digital Endpoints for Clinical Trials: A Structured Approach Toward Fit-for-Purpose Validation,' *Pharmacol Rev*, vol. 72, no. 4, pp. 899–909, Oct. 2020, doi: 10.1124/pr.120.000028.