Dataset-Driven Voice Biomarker Validation: Advancing Clinical Standards with the Enhanced V3 Framework

José Manuel Ramírez, Laura Docío-Fernández, and Carmen García-Mateo

Multimedia Technologies Group, atlanTTic, Universidade de Vigo, 36310 Pontevedra, Spain
Correspondence: jmramirez@gts.uvigo.es

DOI: *https://doi.org/10.17979/spudc.9788497498913.38*

Abstract: Advancements in information technologies are reshaping the medical field, highlighting the need for effective frameworks to translate research into clinical practice. Voice is a promising biomarker for disease detection, but its use is limited by the lack of rigorous clinical validation methods. Our research aims to bridge this gap by improving the validation process for voice-based health assessments. Building on our clinical trial experience and the V3 framework, we are enhancing it to address the specific challenges of voice biomarkers, focusing on the data lifecycle in trials. Our proposal involves refining the V3 framework and developing an open-source toolkit for evaluating voice-based health datasets, bridging the gap between research and clinical practice through robust validation.

1 Introduction

The rapid evolution of information and communication technologies is reshaping the landscape of medicine as we know it. Devices are becoming increasingly compact and affordable, while algorithms are gaining the power to predict medical outcomes with remarkable accuracy. At the forefront of this transformation shift lies Artificial Intelligence (AI) and Digital Medicine (DM), an interdisciplinary field dedicated to harnessing Digital Technologies (DTs) to advance measurement and intervention in human health. DM brings together diverse stakeholders: spanning engineering, manufacturing, clinical science, data science, bio-statistics, regulatory science, ethics, patient advocacy, and healthcare policy. While this interdisciplinary collaboration enriches the field, it also presents challenges, particularly in standardizing terminology and defining best practices Commission et al. (2021, 2024). These challenges have hindered the healthcare industry from fully capitalizing on the potential of DTs.

In this interdisciplinary landscape, digital voice-based measures (DiVMe) emerges as a potent biometric data, offering deep insights into individuals' identities, cultural backgrounds, and emotional states. Its intricate connections to physiological functions and cognitive processes position it as a promising source of biomarkers, providing invaluable insights into brain health. However, the rigorous evaluation of Biometric Monitoring Technologies (BioMeTs), for voice-based biomarkers (VoiBio) or any other biometric data, remains a pressing challenge, hindering their widespread adoption and utilization. In European Union (EU) despite the European Commission's efforts to advance AI and DM in healthcare, adoption is slow and varies across EU countries. The European Parliamentary Research Service reveals on Commission

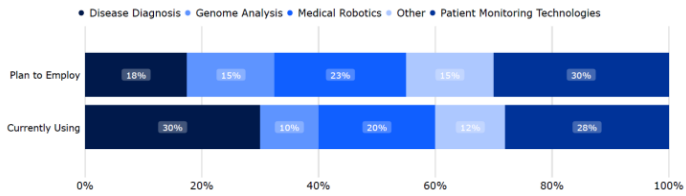


Figure 1: Adoption of AI Applications by Hospitals and Healthcare Service Providers in the EU (taken from Commission et al. (2021)).

et al. (2021), that most healthcare organizations are open to adopting AI, but its use is currently limited to specific departments and applications. The lack of trust in AI-driven decisions and integration challenges are hindering wider adoption. Figure 1 shows some data about the AI adoption by hospitals and healthcare services providers on EU.

The impact of DTs, such as BioMeTs**,** on traditional, analog medical practices has been profound. As DTs advance, the need for regulatory frameworks that can adapt to their unique characteristics becomes increasingly evident. Regulatory bodies, such as the EU´s Directives of 1990 Commission (1990), 1993 Commission (1993), and 2002-2017 Commission (2027), the International Medical Device Regulators Forum in 2013 Group (2013), and the various IEEE Standards (2017) and ISO standards for Standardization (2018), underscore the importance of updating regulations in response to technological progress. These legal frameworks aim to uphold safety, efficacy, and ethical standards in the swiftly evolving realm of DM. Yet, a more holistic approach is required to navigate the complexities of DM effectively.

To address this critical gap, we turn to the V3 framework, a comprehensive methodology developed by Goldsack et al. Robin et al. (2020) for the clinical validation of BioMeTs. The V3 framework encompasses three essential stages: verification, analytical validation, and clinical validation. While V3 provides a valuable foundation, it requires adaptation to meet the specific needs of DiVMe and the EU´s legal framework.

To demonstrate the practical utility of V3 framework, we apply it to two real-world case studies: COSWARA Bhattacharya et al. (2023) and COPERIA Ramírez Sánchez et al. (2023). These studies involve DiVMe from individuals affected by COVID-19, Long-COVID, and healthy controls, providing a diverse dataset for validating VoiBio in different health contexts. This comparative study will allow us to detect those points where V3 falters, allowing us to contribute improvements.

This paper is structured as follows: First, it reviews *Legal and methodological frameworks* in the United States of America (USA) and EU impacting DM. Next, in the *Critical analysis of V3 for VoiBio* section we critically analyze the V3 framework, identifying strengths and gaps, and we also compares the datasets' V3 compliance. The *Conclusion* summarizes findings on clinical validity, assesses V3's utility, and outlines future research directions. It concludes with a *Bibliography*.

2 Legal and methodological frameworks

2.1 USA´s frameworks

The Food and Drug Administration (**FDA**) is the primary regulator of DM technologies, ensuring they are safe and effective. The **Department of Health and Human Services** oversees broader public health initiatives, while the **Office for Civil Rights** focuses on protecting patient health information.

Key USA regulations governing DM technologies include:

- *Food, Drug, and Cosmetic Act* (FD&C Act) U.S. Congress (1934): Gives FDA authority

over medical devices and DM.

- *21st Century Cures Act* Congress (2015): Aims to accelerate healthcare innovation.
- *The Health Insurance Portability and Accountability Act* (HIPAA) Congress (1996): Protects patient data privacy and security.

The FDA has also introduced initiatives like *the Digital Health Innovation Action Plan* FDA (2018) and *Software Pre-Certification Program* FDA (2019) to streamline DM regulation and encourage innovation in this rapidly evolving field. Additionally, the *BEST* (*Biomarkers, EndpointS, and other Tools*) Group (2016) resource provides guidance on the framework for evaluating the clinical validity and utility of DM technologies.

2.2 EU´s frameworks

The EU's regulatory landscape for DM technologies is primarily overseen by the European Medicines Agency (EMA), responsible for the scientific evaluation, supervision and safety monitoring of medicines; the European Commission (EC), responsible for medical device and data protection regulations; and the European Data Protection Board (EDPB), ensuring consistent application of the General Data Protection Regulation (GDPR) Commission (2016). Key EU regulations impacting DM technologies include:

- *Medical Devices Regulation* (MDR) Commission (2017): Governs the safety and performance of medical devices, including software as a medical device (SaMD).
- GDPR Commission (2016): Ensures the protection of personal data and privacy in DM applications.
- *Digital Single Market Strategy* Commission (2024): Aims to ensure fair competition and consumer protection in the digital market.
- *Proposed Artificial Intelligence Act* Commission (2021): Seeks to establish a regulatory framework for safe and transparent AI, including in healthcare.

2.3 Comparison Between the US and EU Regulatory Approaches

The US and EU share key regulatory priorities in medical devices and health data, such as safety, efficacy, and privacy, and both regions are focused on updating frameworks to address AI and digital medicine (DM) technologies. However, their approaches differ in several ways: the EU's more unified regulatory framework (MDR and GDPR) contrasts with the US's fragmented system, with state and federal regulations. The EU emphasizes individual device certification, while the US FDA's Pre-Cert program focuses on software developers. Additionally, the EU's GDPR has stricter data protection standards than the US's HIPAA, and the US tends to support faster market access for innovation, especially through programs like the FDA's Breakthrough Device Designation. Despite these differences, both regions present opportunities and challenges for DM technology manufacturers, with varying regulatory costs and market dynamics. Understanding these regulatory landscapes is crucial for stakeholders aiming to impact medical practice in both regions.

3 Critical analysis of V3 for DiVMe

The increasing use of BioMeTs in healthcare requires a robust evaluation framework that goes beyond regulatory compliance. While legal frameworks like the FDA or the EMA are important, they mainly focus on manufacturers specific claims or risk, not the full capabilities of the technology. This can lead to inconsistent classifications of devices. To ensure safe and effective use, a holistic framework is needed. Goldsack et al. (2020) Goldsack et al. (2020) propose a comprehensive framework (**V3**) to standardize BioMeT evaluation, focusing on **Verification**,

Analytical Validation, and **Clinical Validation**, regardless of regulations or intended use. This approach promotes transparency and informed decision-making in the adoption of these technologies.

In the context of VoiBio, **Verification** entails assessing the quality of DiVMe and scrutinizing the impact of hardware and recording conditions on recording integrity Robin et al. (2020). **Analytical Validation** focuses on verifying the accuracy and reliability of data processing and computed measures, including understanding factors like test-retest reliability and demographic variability Robin et al. (2020). Finally, **Clinical Validation** aims to confirm the correspondence of measures to clinical outcomes such as diagnosis, disease progression, or treatment response Robin et al. (2020). The V3 framework, while a significant step towards standardizing BioMeT evaluation, presents both advantages and shortcomings.

The V3 framework offers several advantages. First, it provides a standardized evaluation process for BioMeTs, ensuring that they meet rigorous quality, safety, and efficacy standards, which facilitates their integration into clinical practice. Additionally, V3 emphasizes clear terminology and standardized procedures, improving communication and collaboration among various stakeholders. However, the framework also has some shortcomings. It lacks detailed guidance on metadata annotations, which is crucial when the intended use or target population of BioMeTs changes, or in cases involving complex symptoms. Without a clear annotation schema, there is a risk of data misinterpretation. Moreover, while V3 touches on ethical considerations, it does not adequately address specific challenges such as data privacy, informed consent, and algorithmic bias, especially in AI-based digital technologies. Lastly, V3 does not provide sufficient detail on how to maintain trust in evolving technologies, highlighting the need for continuous evaluation and feedback to ensure reliability and transparency over time.

3.1 Cases of studies: COSWARA and COPERIA

Real-world medical practice relies heavily on accurate patient measurements and diagnoses. Both regulatory and methodological frameworks aim to ensure a robust and trustworthy process. To this end, we propose analyzing two distinct datasets: COSWARA Bhattacharya et al. (2023), focusing on COVID-19 detection, and COPERIA, emphasizing Post-Acute Sequelae of COVID-19 (PASC, also known as Long-COVID). These datasets, encompassing respiratory sounds from both healthy and affected individuals, offer a unique opportunity for evaluating BioMeTs under the V3 framework.

The COSWARA dataset, developed by the Indian Institute of Science Bangalore, is a diverse collection of respiratory sounds (breathing, coughing, speech) and associated metadata (demographics, health conditions, COVID-19 test results) from 2635 individuals. Its meticulous manual annotation of audio quality makes it a valuable benchmark for evaluating automatic COVID-19 detection systems.

The COPERIA dataset, a collaborative effort by SERGAS, Bahia Software, and Universidade de Vigo, comprises respiratory sounds (coughing and sustained phonation of vowel /a/) and metadata from 138 individuals recorded under two conditions: after physical stress and at rest. COPERIA stands out for its inclusion of clinical diagnoses of Long-COVID and participant medical histories.

V3: Verification

Goldsack et al. (2020) Goldsack et al. (2020) describe verification as a bench evaluation to ensure sensor technologies capture data with minimum accuracy and precision against a reference standard. Both COSWARA and COPERIA datasets use general-purpose devices like cell phones or tablets for collecting respiratory sounds, but their verification processes differ. COSWARA's strength is in its diverse real-world recording environments, though this introduces variability in audio quality, which is mitigated through manual annotation to ensure only high-quality recordings are analyzed. COPERIA, on the other hand, uses a standardized device and con-

trolled environment, reducing variability and aligning with V3's focus on acceptable devices and conditions. Verification in both datasets could be improved with quantitative comparisons of recording quality, such as signal-to-noise ratio (SNR) analysis, and COSWARA would benefit from a more detailed manual annotation protocol. Standardized criteria, like a 50 dBA noise threshold Robin et al. (2020), could also help improve audio quality consistency in future studies.

V3: Analytical Validation

Analytical validation, the second step in the V3 framework, assesses the algorithms that convert raw data into meaningful physiological or behavioral metrics. For both COSWARA and COPERIA, this involves evaluating algorithms that extract features from respiratory sounds and link them to clinical outcomes.

COSWARA uses rich metadata, including COVID-19 test results, to validate its algorithms, though the self-reported nature of diagnoses introduces some uncertainty. Despite this, the dataset has successfully supported AI algorithm development, as shown in the DICOVA challenge, although a detailed statistical analysis is needed. COSWARA also performed a thorough bias analysis across subgroups (e.g., gender, age, vaccination status).

COPERIA benefits from reliable ground truth data, including clinical PASC diagnoses, and its controlled recording environment reduces variability. The experimental design, which includes both rest and post-physical exertion recordings, strengthens algorithm robustness under different conditions. However, a small sample size and potential subjectivity in PASC diagnoses may limit generalizability.

V3: Clinical Validation

Clinical validation, the final step in the V3 framework, assesses whether a BioMeT can accurately identify or predict a meaningful clinical state.

COSWARA leverages diverse demographics to evaluate COVID-19 detection algorithms, but the self-reported nature of COVID-19 diagnoses and the lack of standardized testing protocols across contributors might introduce variability and limit the robustness of clinical validation.

COPERIA focuses on the evolving condition of PASC, using clinical diagnoses and detailed medical histories as a reliable reference. Its controlled environment and specific demographic (middle-aged women) while smaller in scale enhance clinical applicability, aligning with the V3 framework's emphasis on context-specific evaluation.

4 Conclusions

In conclusion, the V3 framework for BioMeT development and validation can benefit from insights gained from the COSWARA and COPERIA case studies. While COSWARA excels in analytical validation with its diverse real-world data and COPERIA provides a controlled environment for clinical validation, both datasets highlight areas for improvement in the V3 framework.

To enhance the framework's utility, it is recommended to incorporate a data sheet for each dataset, clearly documenting the scope, intended use cases, and any limitations regarding data applicability. Additionally, integrating ethical consideration questions into the dataset documentation would address potential concerns related to data privacy and consent. Lastly, implementing distribution metrics to assess and mitigate biases—such as gender bias—would ensure that findings are more representative and equitable across populations. These enhancements would strengthen the V3 framework's overall effectiveness in guiding BioMeT evaluation and adoption.

5 Future Work

We propose developing a comprehensive toolkit for the automatic evaluation of datasets in clinical use, focusing on data integrity and quality in digital medicine research. This toolkit will automate checks for data integrity, utilizing statistical measures such as p-values to assess the representativeness of subsets, ensuring researchers understand the limitations of their data.

Additionally, the toolkit will include guidelines for evaluating demographic distribution metrics to identify and mitigate biases, promoting fairness and equity in clinical research. With a user-friendly interface and comprehensive documentation, this toolkit will empower researchers to assess dataset integrity and address biases effectively, ultimately enhancing the quality and reliability of clinical outcomes.

Bibliography

D. Bhattacharya, N. K. Sharma, D. Dutta, S. R. Chetupalli, P. Mote, S. Ganapathy, C. Chandrakiran, S. Nori, K. Suhail, S. Gonuguntla, et al. Coswara: A respiratory sounds and symptoms dataset for remote screening of sars-cov-2 infection. *Scientific Data*, 10(1):397, 2023.

E. Commission. Directiva 90/385/cee del consejo, de 20 de junio de 1990, relativa a la aproximación de las legislaciones de los estados miembros sobre los productos sanitarios implantables activos, Jul 1990.

E. Commission. Directiva 93/42/cee del consejo, de 14 de junio de 1993, relativa a los productos sanitarios, Jun 1993.

E. Commission. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance), May 2016.

E. Commission. Regulation (eu) 2017/745 of the european parliament and of the council of 5 april 2017 on medical devices, amending directive 2001/83/ec, regulation (ec) no 178/2002 and regulation (ec) no 1223/2009 and repealing council directives 90/385/eec and 93/42/eec, May 2017.

E. Commission. Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, 2021.

E. Commission. Regulation (eu) 2024/903 of the european parliament and of the council of 13 march 2024 laying down measures for a high level of public sector interoperability across the union (interoperable europe act), Mar 2024.

E. Commission. Regulation (eu) 2017/745 of the european parliament and of the council of 5 april 2017 on medical devices, amending directive 2001/83/ec, regulation (ec) no 178/2002 and regulation (ec) no 1223/2009 and repealing council directives 90/385/eec and 93/42/eec, Apr 2027.

E. Commission, C. Directorate-General for Communications Networks, and Technology. *Study on eHealth, interoperability of health data and artificial intelligence for health and care in the European Union – Final study report. Lot 2, Artificial Intelligence for health and care in the EU*. Publications Office of the European Union, 2021.

E. Commission, C. Directorate-General for Communications Networks, Technology, M. Page, R. Winkel, A. Behrooz, and R. Bussink. *2024 digital decade ehealth indicator study – Final report*. Publications Office of the European Union, 2024.

U. Congress. H.r.3103 - 104th congress (1995-1996): Health insurance portability and accountability act of 1996, 8 1996.

U. Congress. H.r. 6 - 114th congress (2015-2016): 21st century cures act, 7 2015.

FDA. Digital health innovation action plan, Mar 2018. URL *https://www.fda.gov/media/106331/download*.

FDA. Software precertifica- tion program, Mar 2019. URL *https://www.fda.gov/media/119722/download*.

I. O. for Standardization. Medical devices - quality management systems - requirements for regulatory purposes (iso 13485:2016). (consolidated version). Standard, International Organization for Standardization, Jul 2018.

J. C. Goldsack, A. Coravos, J. P. Bakker, B. Bent, A. V. Dowling, C. Fitzer-Attas, A. Godfrey, J. G. Godino, N. Gujar, E. Izmailova, et al. Verification, analytical validation, and clinical validation (v3): the foundation of determining fit-for-purpose for biometric monitoring technologies (biomets). *npj digital Medicine*, 3(1):55, 2020.

F.-N. B. W. Group. *BEST* (*Biomarkers, EndpointS, and other Tools*) *Resource* [*Internet*]. Food and Drug Administration (US), Silver Spring (MD) and National Institutes of Health (US), Bethesda (MD), 2016.

I. S. W. Group. Software as a medical device (samd): Key definitions. Tecnical report, International Medical Device Regulation Forum, Dec 2013.

J. Ramírez Sánchez, L. Docío-Fernández, C. García-Mateo, M. Bustillo-Casado, and A. García-Caballero. Identifying patients with long covid: study of the effectiveness of voice signal analysis and machine learning, 7 2023.

J. Robin, J. E. Harrison, L. D. Kaufman, F. Rudzicz, W. Simpson, and M. Yancheva. Evaluation of speech-based digital biomarkers: review and recommendations. *Digital Biomarkers*, 4(3): 99–108, 2020.

I. Standards. Ieee standard for system, software, and hardware verification and validation. *IEEE Std 1012-2016* (*Revision of IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016/Cor1-2017*), pages 1–260, 2017.

U.S. Congress. United States Code: Federal Food, Drug, and Cosmetic Act, 21 U.S.C. §§ 301-392 Suppl. 5. [Periodical], 1934. URL *https://www.loc.gov/item/uscode1934-006021009/*. Retrieved from the Library of Congress.