Clinical Neurophysiology 176 (2025) 2110777

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Review

The GREENBEAN checklist for reporting studies evaluating the effectiveness of EEG-based biomarkers

Joshua B. Ewen [a,*], Claudio Babiloni [b,c], Gary S. Collins [d], Lauren E. Ethridge [e,f], Jean Gotman [g], Akio Ikeda [h], Philippa J. Karoly [i], William Z. Potter [j], Stephan Rampp [k,l], Margitta Seeck [m], Sándor Beniczky [n]

[a] Department of Pediatrics, Ann and Robert H. Lurie Children's Hospital of Chicago and Northwestern Feinberg School of Medicine, Chicago, IL, USA
[b] Department of Physiology and Pharmacology "Vittorio Erspamer", Sapienza University of Rome, Rome, Italy
[c] Hospital San Raffaele Cassino, Cassino (FR), Italy
[d] UK EQUATOR Centre, University of Oxford, Oxford, UK
[e] Department of Psychology, University of Oklahoma, Norman, OK, USA
[f] Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[g] Montreal Neurological Institute-Hospital, McGill University, Montreal, QC, Canada
[h] Department of Epilepsy, Movement Disorders and Physiology, Kyoto University Graduate School of Medicine, Kyoto, Japan
[i] Biomedical Engineering, University of Melbourne, Melbourne, Victoria, Australia
[j] Independent Expert, Bethesda, MD, USA
[k] Department of Neurosurgery, Department of Neuroradiology, University Hospital Erlangen, Erlangen, Germany
[l] Department of Neuroradiology, University Hospital Halle (Saale), Halle (Saale), Germany
[m] Neurocenter, Geneva University, Geneva, Switzerland
[n] Aarhus University and Danish Epilepsy Center, Dianalund, Denmark

ARTICLE INFO

Keywords:
EEG
Biomarker validation
Reporting standard
EQUATOR
Bias
Psychometrics

ABSTRACT

Advances in digital technology, signal analysis, and data science have led to a rapid increase in papers reporting EEG-based biomarkers. However, wide heterogeneity in study design and reporting poses challenges in assessing the reliability, validity and utility of these biomarkers. In this evolving field, best practices are sometimes debated but not yet rigorously defined, and the appropriate next step is to ensure that validation-focused research manuscripts report key methodological factors that are known or suspected to influence results. To assist authors in designing and reporting validation studies of EEG biomarkers, and to help editors and regulatory bodies evaluate them, an international working group—under the auspices of the International Federation of Clinical Neurophysiology (IFCN) and in collaboration with the EQUATOR Network—developed the Guidelines for Reporting EEG/Neurophysiology Biomarker Evaluation for Application to Neurology and Neuropsychiatry (GREENBEAN). EEG biomarker validation studies are classified into four phases, similarly to therapeutic studies. Phases 1–2 are preliminary and do not constitute formal validation. Phase 3 studies provide compelling evidence of validity, while phase 4 studies assess the clinical utility and generalizability of previously validated biomarkers within real-world settings. We provide detailed definitions for each phase, along with a checklist of items to address and report. A detailed *Explanation and Elaboration* document is included in Supplementary Material with multiple examples of how to design and report EEG biomarker studies. We expect that more transparent reporting regarding experimental design and technical standards will not only enhance short-term biomarker validation efforts but will also enhance methodological research to make future efforts more efficient and effective.

Abbreviations: BEST, Biomarkers, Endpoints and Other Tools; COU, Context of Use; EQUATOR, Enhancing the QUAlity Of health Research [Network]; FDA, [US] Food and Drug Administration; GREENBEAN, Guidelines for Reporting EEG/Neurophysiology Biomarker Evaluation for Application to Neurology and Neuropsychiatry; IFCN, International Federation of Clinical Neurophysiology; ISCTM, The International Society for CNS Clinical Trials and Methodology; NIH, [US] National Institutes of Science; STARD, Standards for Reporting Diagnostic accuracy studies.

* Corresponding author at: Division of Developmental & Behavioral Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, 225 E. Chicago Ave, Box 119, Chicago, IL 60611-2605, USA.
E-mail address: jewen@luriechildrens.org (J.B. Ewen).

https://doi.org/10.1016/j.clinph.2025.2110777
Accepted 3 June 2025
Available online 6 June 2025

1. Introduction

Although clinical electroencephalography (EEG), polysomnography and intraoperative monitoring have been standard clinical tools in Neurology for almost a century, recent years have seen many efforts to import EEG-based techniques from research applications to solve a far wider range of clinical needs in Neurology and Neuropsychiatry (Sahin et al., 2018, Ewen and Levin, 2022). First developed as a technology to evaluate human brain physiology by Berger in the late 1920′s (Berger, 1969), EEG quickly became the *sine qua non* clinical tool for epilepsy (Gibbs and Gibbs, 1964, Rossini et al., 2025), and it is currently used in numerous hospitals and neurological practices in economically developed areas, though it may still be out of reach for resource-limited regions.

Certain EEG approaches also developed independently from clinical EEG as research tools for the assessment of psychological and, later, neurobiological processes (Ewen and Levin, 2022). EEG has been successful as a tool in basic science and clinical research (Luck, 2014, Cavanagh, 2019), but efforts in prior decades to translate these approaches to clinical care were often poorly validated and highly controversial within professional medical societies (American Psychiatric Association Task Force on Quantitative Electrophysiological Assessment, 1991, Nuwer, 1997, Ewen, 2016).

Yet as technology and our psychological and neuroscientific knowledge base have evolved, new vistas are open to solve contemporary needs in Neurology and Neuropsychiatry (Levin and Ewen, 2022). EEG-based tools in this vein are typically referred to under the rubric of "biomarkers," which the US National Institutes of Health (NIH) defines as "characteristics that can be objectively measured and used as an indicator of normal biological processes, disease processes, or pharmacologic responses to a therapy" (FDA-NIH Biomarker Working Group, 2021) (Table 1).

Within epilepsy, these needs include diagnosis and classification of epilepsy (including neuroanatomic localization), prognostication and monitoring of therapeutic response. And within Neurodevelopmental Disabilities, Behavioral Neurology and Neuropsychiatry, these needs include techniques that overcome critical workforce shortages or decrease inter-rater subjectivity; have enhanced sensitivity, particularly to fine degrees of change; predict natural-history outcomes or responsiveness to particular interventions, e.g., by identifying distinct biological mechanisms that underlie otherwise indistinguishable behavioral phenotypes (Ewen et al., 2021); and/or parse aspects of cognition and behavior that cannot be disambiguated by traditional (e.g., pen-and-paper or psychophysical) methods, such as arousal, attention, comprehension, effort or motor function, or pathogenic versus compensatory mechanisms.

Table 1
Definitions. Definitions of "biomarker" and "COU" are taken from the FDA-NIH Biomarker Working Group. Some definitions are taken from (Bossuyt et al., 2015).

Term	Definition
Biomarker	Characteristics that can be objectively measured and used as an indicator of normal biological processes, disease processes, or pharmacologic responses to a therapy
Context of Use (COU)	A concise description of the biomarker's specified use (e.g., prognostic, diagnostic, monitoring)
Index Text	The test under evaluation (here, the EEG-based biomarker)
Reference Standard	The method used within a validation study for establishing an outcome or the presence or absence of the target condition; optimally, this would be the best available method. A gold standard would be an error-free reference standard.
Standard-of-Care Test	The current test routinely used in the clinic (or in clinical trials) to report on the outcome or diagnosis of interest. This method is not necessarily equivalent to the Reference Standard. For example, the Wada test can be considered the standard of care for the prediction of memory outcomes in surgical epilepsy patients. The Reference Standard within a predictive biomarker validation study may be the actual longitudinal memory outcomes.

While an expanding toolkit and knowledge basis are critical, they are not sufficient to demonstrate how candidate techniques perform. Thought leaders in EEG have expressed concern that novel EEG biomarker validation attempts suffer from "a lack of adherence to agreed standards and protocols for clinical and scientific practice" (Mushtaq et al., 2024). Developing and validating these techniques requires adaptation of psychometric instrument-assessment methodology (Rudin, 2007, Nunnally and Bernstein, 2010, Ewen and Beniczky, 2018, Ewen et al., 2019). Moreover, evidence-based technical standards for these biomarkers are not yet clear, except in traditional usage around the diagnosis and management of epilepsy (Sinha et al., 2016, Peltola et al., 2023). The goal of these reporting guidelines is to steer investigators toward appropriate validation-study requirements and toward providing sufficient detail about technical practices to enhance reliability and bootstrap research on technical best practices. Both goals help minimize the influence of methodological bias.

Reporting guidelines already exist for validation studies of diagnostic instruments, regardless of modality (Bossuyt et al., 2015). However, we note that EEG-based biomarkers are often called on to perform clinical tasks beyond diagnosis. These clinical tasks have been formalized as "contexts-of-use" (COU) under the US Food and Drug Administration (FDA)/NIH "Biomarkers, Endpoints and Other Tools (BEST)" approaches and include prognostication (of natural history), prediction (of future response to an intervention), monitoring and response (concurrent monitoring of response to an intervention), and risk and safety (prediction of adverse response to an intervention) (Table 2) (FDA-NIH Biomarker Working Group, 2021). To encompass not only the specific technological characteristics of EEG but also the range of COUs for which EEG-based biomarker are used, we developed a new guideline rather than an extension to the "Standards for Reporting Diagnostic accuracy studies" (STARD; Bossuyt et al., 2015) or other biomarker development and evaluation frameworks, under the auspices of the International Federation of Clinical Neurophysiology (IFCN) and in collaboration with the EQUATOR (Enhancing the QUAlity Of health Research) Network. Titled "Guidelines for Reporting EEG/Neurophysiological Biomarker Evaluation for Application to Neurology and Neuropsychiatry" (GREENBEAN), these reporting guidelines are intended to fill a scope beyond the diagnosis-only focus of STARD (Bossuyt et al., 2015), the specific-condition-focus of other guidelines (Webb et al., 2015), specific-EEG-analysis-type focus of yet others (Picton et al., 2000), and the digital health focus of others (Goldsack et al., 2020). As a concrete example, classical biomarker concepts like analytical validation are uniquely challenging when a reference standard cannot be easily ascertained for a novel method, as in the case of signals derived from brain electrical activity. Thus, new steps for biomarker development and evaluation should be considered in the context of EEG to reflect these challenges.

We recognize that these guidelines encourage the reporting of a good deal of information and feel that this information is necessary until such time as best practices and standards crystallize. In the meanwhile, we suggest that journal editors should bear in mind the information needed to replicate biomarker-validation studies in the laboratory and, eventually, in the clinic, and should assign word limits accordingly.

2. Guideline development methods

The IFCN Executive Committee convened a working group to provide reporting guidelines around EEG-based biomarker validation, in order to increase the quality of biomarker studies and validation reports. Working-group members were recruited based on their experience in a wide range of EEG applications, including epilepsy ("standard" clinical EEG applications, plus novel signal-analysis approaches) and EEG-based biomarker development for neurocognitive disorders (cognitive task development, multisite study implementation, and technical expertise in

Table 2
Biomarker Contexts of Use (COU) (FDA-NIH Biomarker Working Group, 2021).

Context of Use	Description
Diagnostic	A biomarker used to detect or confirm presence of a disease or condition of interest or to identify individualswith a subtype of the disease
Monitoring	A biomarker measured repeatedly for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent
Response	A biomarker used to show that a biological response, potentially beneficial or harmful, has occurred in an individual who has been exposed to a medical product or an environmental agent • *Pharmacodynamic biomarker:* A response biomarker that indicates biologic activity of a medical product or environmental agent without necessarily drawing conclusions about efficacy or disease outcome or necessarily linking this activity to an established mechanism of action. Potential uses of a pharmacodynamic biomarker include establishing proof-of-concept, assisting in dose selection or measuring a response to medical products or environmental agents, including the use as a measure of potential harm. In some cases, such measures may be secondary endpoints in clinical trials and may be described in labeling. • *Surrogate endpoint biomarker:* A response biomarker that is an endpoint used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientificevidence.
Predictive	A biomarker used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent
Prognostic	A biomarker used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest
Safety	A biomarker measured before or after an exposure to a medical product or an environmental agent to indicate the likelihood, presence, or extent of toxicity as an adverse effect
Susceptibility/ Risk	A biomarker that indicates the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or the medical condition

recording in special populations), as well as knowledge of regulatory processes and reporting standard development. The group consisted of experts from Europe, North America, Asia and Oceania. The basic format proceeded along the lines of prior efforts (Moher et al., 2010, Abi-Dargham and Horga, 2016, Woo et al., 2017).

At a series of introductory online meetings, working-group members were provided several relevant publications on technical standards of EEG and publication standards (McShane et al., 2005, Schunemann et al., 2008, Whiting et al., 2011, Bossuyt et al., 2015, Ewen and Beniczky, 2018, Ewen et al., 2019). Members were then asked to submit items for consideration. Additional items were collected from other guidelines (Picton et al., 2000, Webb et al., 2015, Webb et al., 2019). A Delphi process (Taylor, 2020) was then initiated to vote on inclusion of items amongst working-group members, using Welphi web-based software (DecisionEyes, Lisbon, Portugal). Pre-set criteria were $\geq$67 % "yes" for automatic inclusion and $<$33 % "yes" for automatic rejection. Members were encouraged to submit comments in response to each item. Two rounds were conducted, with items being revised in between. Following the Delphi rounds, over 80 items had been approved and included. The two working-group leaders combined items that were effectively redundant or that were categorically similar such that only 34 unique items remained. The revised checklist and draft of the Explanation & Elaboration text were sent to the entire working group for comments and edits. These edits were revised by the leaders and sent, along with the primary document, to be approved by the whole working group. These were then sent to the Executive Committee of the IFCN for approval and then distribution for public comment, including from authors of the publications we cited as examples. We received feedback from 27 experts, including formal feedback from The International Society for CNS Clinical Trials and Methodology (ISCTM) (see acknowledgements). The document was then revised to incorporate suggestions aligned with the guideline's objectives. Further explanations of each item and illustrative examples are published in an extensive ***Explanation and Elaboration document***, contained in Supplemental Information. *These elaborations specify in detail the information that should be included for each item.* This document contains many examples of existing biomarker studies, explains how to report studies (e.g. what the title, introduction and methods should include), how the index test should be described including technical aspects and Context of Use, how to select reference standard, guidelines on subject selection and statistical analyses. For presentation of Results, the document explains how to present the flow of participants and factors that could affect the results as well as descriptions of what the added value of the index test is.

2.1. Phases of biomarker validation

The committee recognized that biomarker validation seems to evolve in steps, and different subsets of reporting-guideline items would be relevant for studies on different steps of this process. We therefore undertook to specify phases of validation as part of the Delphi process, building on a prior rubric (Ewen et al., 2019). Our rubric is similar to but differs slightly in content from other, similar efforts (Abi-Dargham and Horga, 2016).

Phase 1: Biomarker discovery. These are exploratory studies that report a statistical relationship between the EEG metric and some aspect of the clinical state or diagnosis. They may include case-control comparisons, two-group comparisons, correlation between a physiological (EEG) measure and a clinical variable, data-driven cluster identification, or identification of an EEG measure in which a clinical group is in the tail of normative distribution. Particularly for prognostic, predictive, monitoring and risk COUs, these studies can also include longitudinal data collection in which many EEG variables are explored. These studies are not formally validation studies.

Phase 2: Biomarker preliminary validation. Preliminary validation seeks to begin to understand how a particular EEG-based candidate biomarker performs at the individual-patient level, within its defined COU. It is often technically simpler than full validation (Phase 3). Phase 2 studies include research encompassing any or several of the following goals: to establish optimal thresholds (e.g., operational point on a Receiver Operator Curve) in cases when a binary scale of measurement is employed, use cross-validation or a small/single-center dataset to estimate within-sample accuracy/predictive performance, or establish test–retest measurement of reliability. As test–retest reliability sets a ceiling for quantitative validity, reliability data may be particularly useful in determining whether to pursue further validation studies.

Phase 3: Biomarker validation. These studies provide compelling evidence for accuracy/predictive value of the biomarker. They often use independent "test" datasets (as opposed to Phase 2, in which training and test typically occur in a single dataset) to estimate out-of-sample accuracy measures (including sensitivity/specificity, positive and negative predictive values for binary scale-of-measurement; confidence intervals or R^2 for continuous scale-of-measurement). A validated biomarker is appropriate for use within the constrained context under which it was validated. Phase 3 studies may build on prior Phase 2 studies that have already pre-identified analysis pipelines, relevant parameters and thresholds, or they may include both training sets and independent test sets. Requirements for Phase 3 studies are highlighted in Table 3.

Phase 4A: Clinical utility. Phase 4A studies build on biomarkers validated in a particular COU within Phase 3, by demonstrating that use of the biomarker improves patient outcomes, accounting for its impact at the individual patient level, rather than group differences.

Phase 4B: Independence from specific analytic methods. Phase 4B studies build on biomarkers validated in Phase 3 by demonstrating

that the biomarker works across multiple different analysis pipelines, environments or equipment. Such studies may also demonstrate independence from a particular patient state (e.g., wake-sleep or task) in which the biomarker was initially validated. This phase is similar in goals to the US FDA Biomarker Qualification Program (Amur et al., 2015).

Phase 4C: External clinical and demographic generalization. This phase takes a biomarker already validated in a Phase 3 study and demonstrates that it also performs to a clinically useful degree within age groups, diagnostic groups, interventions and so on that are non-identical but similar to the age groups, etc., in which it was originally validated.

2.2. Reporting items

The key reporting items, as identified by the working group, are listed in Table 4. They encompass experimental-design considerations typical of clinical-instrument validation studies and technical points specific to EEG. While psychometric experimental design is well established (Nunnally and Bernstein, 2010), the "best practices" for EEG technical procedures across the range of different data collection (i.e., spontaneous vs. task) and signal-processing techniques are still evolving. The data required by these reporting standards will not only allow reviewers, readers and regulators to assess for sources of bias but will also enhance the ability of independent laboratories to replicate the techniques and achieve similar biomarker performance.

The ***Explanation and Elaboration*** document (Supplementary Information) provides multiple examples of how to deal with conceptual and practical issues of defining and reporting EEG biomarkers in the different categories.

2.3. Adapting the reporting requirements to individual studies

While reporting guidelines are, by their nature, prescriptive, we also recognize that not all items will be relevant to a particular study. The items that are helpful to a reader and should be required by an editor will vary by the candidate biomarker's COU and the phase of validation. The items required by a particular study will also differ (Table 5) based on additional factors, including.

- Task-based vs. spontaneous ("resting state")
- Longitudinal vs. cross-sectional data collection
- Prospective vs. retrospective sampling
- Presence vs. absence of an intervention
- Multisite vs. single site (recognizing however that a valid biomarker will need to be able to be used in diverse laboratories and not simply the site that validated it)
- Scale of measurement (e.g., binary, continuous)
- Whether the estimate of the outcome takes into account only the EEG results or other data as well

Table 3
Key aspects of an adequate Phase 3 validation study.

Feature	Commentary
Starting point	Addressing unmet clinical needs within its Context of Use. (Explain the limitations of current "standard of care" methods of assessment and how this is addressed by the index test).
Study design	Prospective (or retrospective but blinded)
Test dataset	Independent and representative Large, multicenterPrevalence of diagnosis/outcome in training and test samples should reflect the population of the intended useSample size sufficient for the target sensitivity and specificity
Index test	Pre-defined model/algorithmPre-defined cut-off values
Reference standard	Adequate choice of reference standard Reliable method for determining the reference standard

Table 4
The GREENBEAN checklist.

Section & Topic	No	Item	Applicability Notes	Page
TITLE OR ABSTRACT				
	1	Identify "biomarker," "EEG." Specify Context of Use and phase of validation		
INTRODUCTION				
	2	State unmet clinical need to be addressed and context of Use. Explain the limitations of current "standard of care" methods of assessment		
	3	Justification for EEG features studied		
METHODS				
Study Design	4	Description of the study design, including any longitudinal data collection and presence or absence of intervention		
Participants	5	Inclusion/exclusion criteria, ascertainment/ recruitment procedures (including prospective/ retrospective design), stratification procedures (if applicable), and relationship among them		
	6	Describe how the data were used (e.g., for training and validation/ test) in the analysis, considering sample size requirements	Phases 2–4 only	
	7	Prevalence of diagnosis/ outcome in training and test samples or distribution of clinical independent variable		
	8	Justification of control group	When different samples are recruited separately	
Test Methods	9	Index test, in sufficient detail to allow replication a. Participant behavioral management, b. hardware and software, including electrode montage, electrode reference, and recording settings, c. recording environment, d. technologist qualifications/ training/certification, e. psychophysical equipment, tasks, order, and instructions for participants, f. calibration of stimulus-delivery devices,		

(*continued on next page*)

Table 4 (*continued*)

Section & Topic	No	Item	Applicability Notes	Page
		g. Data analysis pipeline in sufficient detail to allow replication, including preprocessing of EEG data, data selection/rejection, approaches/criteria/parameters,		
		h. quality control procedures, including staff training and inter-site calibration		
	10	Disclosure of all clinical information or results available to personnel involved in data collection and analysis (both index and reference tests), specifying if they were blinded to certain data		
	11	Pre-specification of EEG dependent variables, analysis pipeline and thresholds, and disclosure of all investigator-set parameters or steps explored/revised after data collection began	Required for test samples of Phases 3–4	
	12	Reference standard, its rationale, and procedures for determining alternative diagnoses/outcomes, in sufficient detail to allow replication		
	13	Criteria for any exclusion of any participants following (attempted) data collection		
Statistical Analysis	14	All other risk, disease and demographic variables analyzed alongside EEG dependent variables		
	15	For multivariable prediction models, details of any variable selection procedures, data-reduction strategies, and other model-building issues	For studies that include multivariable models	
	16	How missing data and outliers were handled and practical reasons for any missing values		
	17	Methods for estimating or comparing measures of diagnostic accuracy, disease progression, etc.		
	18	Details of any correction for multiple comparisons	For Phases 1–2 or the training samples of Phase 3	
RESULTS				
Participants	19	Flow of participants, using a diagram; including details on participants who did not tolerate testing, did not have interpretable data or withdrew from the study (including explanation why they did so)		
	20	All factors relevant to clinical outcome and that can influence the EEG result, including baseline demographic and clinical characteristics of participants, reporting differences between samples (disease/control or training/test)		
	21	Time interval and any clinical interventions between index test and reference standard		
	22	Change in any clinical variables over the course of a longitudinal study		
Test Results	23	a. Estimates of accuracy and their precision (such as confidence intervals)	Phases 3–4	
		b. Results demonstrating the added value of EEG biomarker beyond other data sources known to the clinician		
	24	For training sample, ROC curve by classification threshold	For some Phase 2 and all Phase 3 (training sample) studies that employ binarized outcomes	
	25	Graphical visualization of relevant variability and/or consistency in the EEG output (e.g., ERP grand average with confidence intervals, split-half plots or butterfly plots)		
	26	For samples where the prevalence of relevant independent variables (e.g., diagnosis prevalence) or covariates (e.g., age) are dissimilar to the typical clinical population of interest on whom the biomarker will be employed, estimates of population PPV/NPV		
DISCUSSION				
	27	Study limitations, including sources of potential bias/confounding, statistical uncertainty, and generalizability		
	28	Practical implementation factors, including cost of implementation, training and certification of staff, quality control, stigma, need for special testing environments, etc.		
	29	Feasibility of all tasks for the clinical population		
	30	Anticipated or demonstrated impact on patient outcomes	Requirement for Phase 4A	
	31	Demonstration that biomarkers contribute information beyond other sources of (demographic, clinical/disease)		

(*continued on next page*)

Table 4 (*continued*)

Section & Topic	No	Item	Applicability Notes	Page
		information that will be known by clinician		
OTHER CONSIDERATIONS				
	32	Provide details about the availability of analysis code, software and data		
	33	Sources of funding and other support; role of funders		
	34	Use of generative AI in the production of the manuscript		

2.4. Availability

Open-access publication, including the ***Explanation and Elaboration*** document, in *Clinical Neurophysiology* journal, the IFCN website and the EQUATOR Network website.

3. Future directions

These reporting guidelines merely prescribe the information that reviewers, readers, funders and regulatory agencies should have access to about a particular EEG-based biomarker validation study. They are intentionally not prescriptive about best practices, as the effect of the various data-collection and analysis choices on reproducibility of EEG analysis output is not yet fully known. As one example, unlike known effects of different MRI scanners, the sensitivity to individual types of metrics, be they frequency-based, time-based, localization-based or bivariate (connectivity), to differences in electrode composition or amplifier characteristics is not fully known. We encourage ongoing methodological research to determine which factors have the greatest impact—and which have negligible impact—on various signal analyses (e.g., Trubutschek et al., 2024). As the field matures, we eagerly anticipate authors being able to know that, by following empirically-justified best practices, their studies have minimal exposure to methodological forms of bias. Such best practices will progressively supersede reporting guidelines.

4. Sponsorship

These guidelines were developed under the auspices of the IFCN.

5. Funding sources

JBE's time was supported in part by the Kennedy Krieger Institute Intellectual and Developmental Disabilities Research Center, NIH/NICHD grant number P50HD103538. MS's effort was supported by Swiss National Science Foundation grant CRS115-180365.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: JBE has consulted for Novartis on biomarker development. LEE has consulted for Novartis, Tetra Therapeutics a Shionogi Company, Healx, Ultragenyx, Autifony Therapeutics, and Purposeful on EEG biomarker outcome measurement. AI Belongs to the Department of Epilepsy, Movement Disorders, and Physiology is the Industry-Academia Collaboration Course, being supported by Eisai Co., Ltd., Nihon Kohden Corporation, Otsuka Pharmaceutical Co., UCB Japan Co., Ltd., Sumitomo Pharma, and RICOH Company. MS owns shares CloudsofCare and dEEGtal and received speaker fees from Bial and EISAI.

Table 5
Variations in applicable GREENBEAN items based on study design factors. All items are relevant unless specified in the table. All Title, Abstract and Introduction items are equally relevant to all studies.

	Methods	Results	Discussion
Task-Based vs. Spontaneous	Item #9e is not relevant for spontaneous recordings	All items are relevant	All items are relevant
By Sampling Approach	Items #5, 7 and 8 will be particularly important for studies in which groups are binarized (either retrospectively or prospectively; Item #9 will be particularly important in case-control (retrospective) designs	Item #26 will be particularly important for determining differences when the independent variable is binarized	All items are relevant
Concurrent vs. Longitudinal	Item #4 will highlight any longitudinal data collection	Item #21 would include any time-interval between the index test (EEG biomarker recording) and reference standard, whether due to intentionally longitudinal design or practical delays within a concurrent design; Item #22 irrelevant to concurrent studies	All items are relevant
Presence vs. Absence of an Intervention	For intervention-including studies, Item 4 would describe the intervention	Item #22 would report the effects of the intervention within an intervention-including study	All items are relevant
By Phase	For Phase 1 and 2 studies, Item #11 would be irrelevant; Phase 4 studies presumably do not contain a training sample, therefore Item 18 would be irrelevant	All items are relevant	Item #30 is empirically based for Phase 4A studies and speculative for other phases
Multi-Site vs. Single Site	Item #9h is irrelevant to single-site studies	All items are relevant	All items are relevant
Scale of Measurement	Items #5, 8, 9 are particularly relevant for a binarized independent variable (e.g., two patient groups)	Item #24 is irrelevant for non-binarized variables	All items are relevant
Unimodal vs. Multimodal	Item #15 may be less relevant for studies in which there is only one dependent variable (EEG); for studies without multivariable models, Item #15 is irrelevant	Item #23b is less relevant when only unimodal EEG data is recorded	In studies in which only unimodal EEG data is collected (i.e., without the other clinical data a clinician would have), the explanation of why such data was not collected as a part of the study is critical (see Item #31)

Acknowledgements

The authors would like to thank International Society for CNS Clinical Trials and Methodology (ISCTM), who convened a group led by Kemi Olugemo and Atul Mahableshwarkar, to review and provide comments on behalf of the Society. The group also included Michael Avissar, Kjartan Frisch Herrik, Seth Hopkins, Bruce Kinon, Kristiina Kompus, David Matthews, Dona Murphey, Aneeta Saxena, and Leif Simmatis. The ISCTM has not endorsed the final draft. The authors would also like to thank the community experts who contributed feedback, including Ravindra Arya, Alana Campbell, James Cavanagh, Eric Chin, Mark Hallett, Jeremy Harper, Alexandra Key, Dara Manoach, Yasue Mitsukura, Sebastian Olbrich, Mark Pflieger, Ernest Pedapati, Guillaume Rousselet, Olga Selioutski, Saurabh Sinha, Jeffrey Tenney. The published work does not represent an endorsement by any of these experts, their institutions or the professional societies with which they are involved.

Appendix A. Supplementary data

Supplementary data containing the full Explanation and Elaboration content for this reporting guideline can be found online at https://doi.org/10.1016/j.clinph.2025.2110777.

References

Abi-Dargham, A., Horga, G., 2016. The search for imaging biomarkers in psychiatric disorders. Nat. Med. 22, 1248–1255.

American Psychiatric Association Task Force on Quantitative Electrophysiological Assessment. Quantitative electroencephalography: a report on the present state of computerized EEG techniques. Am J Psychiatry. 1991;148:961-4.

Amur, S.G., Sanyal, S., Chakravarty, A.G., Noone, M.H., Kaiser, J., McCune, S., et al., 2015. Building a roadmap to biomarker qualification: challenges and opportunities. Biomark. Med. 9, 1095–1105.

Berger, H., 1969. On the Electroencephalogram of Man: Twelfth Report. Electroencep. Clin. Neurophysiol. Suppl. 28, 267–287.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L., et al., 2015. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, h5527.

Cavanagh, J.F., 2019. Electrophysiology as a theoretical and methodological hub for the neural sciences. Psychophysiology 56, e13314.

Ewen, J.B., 2016. The eternal promise of EEG-based biomarkers: getting closer? Neurology 87, 2288–2289.

Ewen, J.B., Beniczky, S., 2018. Validating biomarkers and diagnostic tests in clinical neurophysiology: Developing strong experimental designs and recognizing confounds. In: Schomer, D.L., Lopes da Silva, F.H. (Eds.), Niedermeyer's Electroencephalography, 7th ed. Oxford University Press, New York.

Ewen, J.B., Levin, A.R., 2022. Neurobehavioral biomarkers: an EEG family reunion. J. Clin. Neurophysiol. 39, 129–134.

Ewen, J.B., Potter, W.Z., Sweeney, J.A., 2021. Biomarkers and neurobehavioral diagnosis. Biomarkers Neuropsychiatry.

Ewen, J.B., Sweeney, J.A., Potter, W.Z., 2019. Conceptual, regulatory and strategic imperatives in the early days of EEG-based biomarker validation for neurodevelopmental disabilities. Front. Integr. Neurosci. 13, 45.

FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring, MD and Bethesda, MD: Food and Drug Administration (US) and National Institutes of Health (US); 2021.

Gibbs, F.A., Gibbs, E.L., 1964. Atlas of Electroencephalography. Addision-Wesley, Reading, MA.

Goldsack, J.C., Coravos, A., Bakker, J.P., Bent, B., Dowling, A.V., Fitzer-Attas, C., et al., 2020. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit. Med. 3, 55.

Levin, A.R., Ewen, J.B., 2022. Bridges Through the Cloud: Towards Clinical Biomarkers of Cognitive Neurophysiology. J. Clin. Neurophysiol. 39, 99–100.

Luck, S., 2014. An Introduction to the Event-Related Potential Technique, second ed. MIT Press (A Bradford Book), Cambridge.

McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., Clark, G.M., et al., 2005. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br. J. Cancer 93, 387–391.

Moher, D., Schulz, K.F., Simera, I., Altman, D.G., 2010. Guidance for developers of health research reporting guidelines. PLoS Med. 7, e1000217.

Mushtaq, F., Welke, D., Gallagher, A., Pavlov, Y.G., Kouara, L., Bosch-Bayard, J., et al., 2024. One hundred years of EEG for brain and behaviour research. Nat. Hum. Behav. 8, 1437–1443.

Nunnally, J.C., Bernstein, I.H., 2010. Psychometric Theory. Tata McGraw-Hill.

Nuwer, M., 1997. Assessment of digital EEG, quantitative EEG, and EEG brain mapping: report of the American Academy of Neurology and the American Clinical Neurophysiology Society. Neurology 49, 277–292.

Peltola, M.E., Leitinger, M., Halford, J.J., Vinayan, K.P., Kobayashi, K., Pressler, R.M., et al., 2023. Routine and sleep EEG: minimum recording standards of the international federation of clinical neurophysiology and the international league against epilepsy. Epilepsia 64, 602–618.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson Jr, R., et al., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37, 127–152.

Rossini, P.M., Cole, J., Paulus, W., Ziemann, U., Chen, R., 2025. 1924-2024: First centennial of EEG. Clin. Neurophysiol. 170, 132–135.

Rudin, M., 2007. Imaging readouts as biomarkers or surrogate parameters for the assessment of therapeutic interventions. Eur. Radiol. 17, 2441–2457.

Sahin, M., Jones, S.R., Sweeney, J.A., Berry-Kravis, E., Connors, B.W., Ewen, J.B., et al., 2018. Discovering translational biomarkers in neurodevelopmental disorders. Nat. Rev. Drug Discov.

Schunemann, H.J., Oxman, A.D., Brozek, J., Glasziou, P., Jaeschke, R., Vist, G.E., et al., 2008. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 336, 1106–1110.

Sinha, S.R., Sullivan, L., Sabau, D., San-Juan, D., Dombrowski, K.E., Halford, J.J., et al., 2016. American clinical neurophysiology society guideline 1: minimum technical requirements for performing clinical electroencephalography. J. Clin. Neurophysiol. 33, 303–307.

Taylor, E., 2020. We Agree, don't we? The Delphi method for health environments research. HERD 13, 11–23.

Trubutschek, D., Yang, Y.F., Gianelli, C., Cesnaite, E., Fischer, N.L., Vinding, M.C., et al., 2024. EEGManyPipelines: a large-scale, grassroots multi-analyst study of electroencephalography analysis practices in the wild. J. Cogn. Neurosci. 36, 217–224.

Webb, S.J., Bernier, R., Henderson, H.A., Johnson, M.H., Jones, E.J., Lerner, M.D., et al., 2015. Guidelines and best practices for electrophysiological data collection, analysis and reporting in autism. J. Autism Dev. Disord. 45, 425–443.

Webb, S.J., Shic, F., Murias, M., Sugar, C.A., Naples, A.J., Barney, E., et al., 2019. Biomarker acquisition and quality control for multi-site studies: the autism biomarkers consortium for clinical trials. Front. Integr. Neurosci. 13, 71.

Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., et al., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20, 365–377.