NeuroImage 245 (2021) 118721

Contents lists available at ScienceDirect

NeuroImage

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

Towards ARTEM-IS: Design guidelines for evidence-based EEG methodology reporting tools

Suzy J Styles[a,b,c,*], Vanja Ković[d], Han Ke[a], Anđela Šoškić[d,e,*]

[a] Psychology, Nanyang Technological University, Singapore
[b] Centre for Research and Development in Learning, Nanyang Technological University, Singapore
[c] Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore
[d] Laboratory for Neurocognition and Applied Cognition, Faculty of Philosophy, University of Belgrade, Singapore
[e] Teacher Education Faculty, University of Belgrade, Singapore

ARTICLE INFO

Keywords:
EEG
Neuroscience
Methodology
Human error
International standards
ARTEM-IS

ABSTRACT

As the number of EEG papers increases, so too do the number of guidelines for how to report what has been done. However, current guidelines and checklists appear to have limited adoption, as systematic reviews have shown the journal article format is highly prone to errors, ambiguities and omissions of methodological details. This is a problem for transparency in the scientific record, along with reproducibility and metascience. Following lessons learned in the high complexity fields of aviation and surgery, we conclude that new tools are needed to overcome the limitations of written methodology descriptions, and that these tools should be developed through community consultation to ensure that they have the most utility for EEG stakeholders. As a first step in tool development, we present the ARTEM-IS Statement describing what action will be needed to create an Agreed Reporting Template for Electroencephalography Methodology - International Standard (ARTEM-IS), along with ARTEM-IS Design Guidelines for developing tools that use an evidence-based approach to error reduction. We first launched the statement at the LiveMEEG conference in 2020 along with a draft of an ARTEM-IS template for public consultation. Members of the EEG community are invited to join this collective effort to create evidence-based tools that will help make the process of reporting methodology intuitive to complete and foolproof by design.

1. Introduction

In any research context where multiple decisions must be made about how to handle, process and analyse data, a researcher has a high number of 'degrees of freedom' in their methodology. This phenomenon is also known as the 'garden of forking paths' (Gelman and Loken, 2013). In the field of EEG, and neuroscience in general, the number of different methodological decision points is extremely high, massively inflating the possible 'paths' a researcher might take through the garden. For example, in one preregistered study, Šoškić et al. (2019) have shown that implementing only 2-4 'settings' for each of 11 'steps' in an EEG pre-processing and analysis pipeline results in no less than 864 possible 'paths' through the garden.

With so many methodological decisions to be reported, omissions and ambiguities are likely to occur in the process of reporting, revising, and copy-editing written methodology descriptions in journal articles. For example, in one systematic review, Šoškić and colleagues (2021) evaluated methodology reporting in 132 peer reviewed N400 journal articles published between 1988-2018. The results showed that some methodological items were reported in most, although not all, cases (e.g., sample size, measurement window, offline filter cut-offs, number of trials presented). Others were frequently omitted, or reported with substantial ambiguities (e.g., filter roll-off and slope details, trials remaining after rejections, order of pre-processing operations). In some cases it was possible to infer omitted values from other reporting items (e.g., sample size following exclusions could be reconstructed from the degrees of freedom in statistical analyses). However, other items were not possible to infer or reconstruct (e.g., order of operations). Overall all papers in the systematic review omitted at least some methodological details, and 46% of papers contained ambiguities. These omissions and ambiguities demonstrate a lack of transparency in the current scientific literature.

* Corresponding author at: Psychology, Nanyang Technological University, Singapore.
E-mail addresses: suzy.styles@ntu.edu.sg (S.J. Styles), andjela.soskic@uf.bg.ac.rs (A. Šoškić).

https://doi.org/10.1016/j.neuroimage.2021.118721.
Received 12 June 2021; Received in revised form 11 October 2021; Accepted 3 November 2021
Available online 24 November 2021.

1.1. Consequences for reproducibility and replicability

If methodology reports contain errors, omissions and ambiguities, this prevents a researcher following the precise sequence of steps in subsequent studies that aim to either reanalyse an existing data set (i.e., reproducing results) or run a previously validated study with a new sample of participants (i.e., replicating a study). It could be argued that ambiguities in a methodological pathway may not be pressing if most pathways lead to the same outcome (i.e., if outcomes are robust to changes in methodological decisions). However, studies investigating the robustness of EEG results to methodological variation have repeatedly shown that this is not the case. Following the same path except for one crossroad: Multiple publications have shown that individual decisions and steps in the pipeline have the potential to modify the effect size and study outcomes, as has been shown for sample size (Boudewyn et al., 2018), EEG recording systems (Melnik et al., 2017), electrode impedance (Kappenman and Luck, 2010), filters (Tanner et al., 2015), and statistical analysis (Luck and Gaspelin, 2017). Picking multiple different paths through the garden: In the EEG subdomain of event-related potentials (ERP), two multiverse reanalyses have examined consequences of systematically sampling across the methodological possibility space, and shown that different analysis pathways have consequences on study outcomes such as the difference between conditions, effect size, statistical power, psychometric properties and their association with gender and behaviour (Sandre et al., 2020; Šoškić et al., 2019).

Furthermore, Šoškić et al. (2021) argue that this is not just a theoretical concern for methodologists, but that lack of detail in the published literature has real world outcomes in the field of replication science. In a recent multi-lab study (Nieuwland et al., 2018), which failed to replicate an influential N400 paper by DeLong, Urbach and Kutas (2005), one of the points of discussion about the reasons for failure was adherence to details of the pre-processing pipeline, more specifically, baseline correction (DeLong et al., 2017; Nieuwland et al., 2018). This discussion underscores the importance of providing full methodological reports in order to allow direct replication of procedures, and, consequently, adequate assessment of the replicability of results. This challenge can only become greater as the larger replication projects, such as the EEGManyLabs (Pavlov et al., 2021), are emerging in the field, in line with general movements for greater transparency and replicability in the medical and psychological sciences (Ioannidis et al., 2014; Nosek et al., 2015).

It is important to note that inability to replicate previously published effects is likely to have the greatest impact on researchers new to the field of EEG research, who may not be the beneficiaries of decades long experience in particular methodological traditions, or the knowledge contained in unpublished materials like 'lab handbooks' which may be passed down through generations of scholars within a research team. This means the researchers most impacted by the failure to replicate previous research are likely to be those in new or underfunded research environments, and at early career stages - a sentiment echoed by Bishop (2017) and incorporated into the British Neuroscience Association's 'Credibility in Neuroscience Manifesto' (2019).

1.2. Consequences for metascience

In addition to consequences for individual studies, if methodology reports contain errors, omissions and ambiguities, this can create obstacles for a researcher who wants to combine the results of previous studies into a large-scale review using the tools of meta-science (e.g., meta-analysis, or qualitative synthesis). A PubMed search for the phrase *event related potentials* with a filter for "systematic review" gives 267 findings since 1998, 61% of which have been published since 2018 (on 27th May 2021). This large number demonstrates substantial interest in synthesizing empirical findings across multiple individual studies. As we have shown, the brief verbal descriptions in journal articles obscure similarities and differences between publications. As a result, it is not possible to identify and group papers that have used similar, or even identical processing, or to establish how much of the variability found in the literature can be attributed to methodological choices. Indeed, the errors and ambiguities in reporting observed in the Šoškić et al., systematic review (2021) were so pervasive that our review team were obliged to abandoned a secondary plan to conduct statistical meta-analysis of the sample. Moreover, as methodological documentation is almost exclusively limited to written descriptions, the present formats mean that both search of existing literature based on methodological criteria, and extraction of methodological details for metascience are laborious and/or ineffective.

2. How has the field attempted to improve reporting to date?

Historically, as the field of electrophysiological research has grown, so have calls for standardization of research methods, and methods reporting. As an illustration, Fig. 1 shows the emergence of articles containing the terms 'ERP' and 'Event related potential(s)', along with the publication of notable general guidelines (Donchin et al., 1977; Keil et al., 2014; Pernet et al., 2018a; Picton et al., 2000), subfield guidelines (Duncan et al., 2009; Kappenman and Luck, 2016; Taylor and Baldeweg, 2002), and practice manuals (Handy, 2005; Luck, 2005, 2014), all of which call for increased attention to the specific details in an electrophysiological study, and clear reporting of what has been done. Over the decades, these methodological calls have evolved - The earliest guidelines give detailed verbal descriptions of the steps involved in recording and processing electrophysiological data, along with guidance about how to adequately describe these details in research reports (Donchin et al., 1977; Picton et al., 2000); These early publications were followed by practice manuals providing a more detailed treatment of different methodological elements, including guidance on best practice, as well as reporting (Handy, 2005; Luck, 2005, 2014). More recently, methodological publications in this field have included reporting 'checklists' itemizing different details that should be reported in a research publication, as a strategy for improving compliance with stated reporting guidelines (Keil et al., 2014; Pernet et al., 2018a).

3. Is it getting better?

Despite the advances of the past 45 years, uptake of reporting guidelines remains moderate at best, and there is limited evidence that reporting checklists are broadly used. For example, Clayson et al. (2019) compared articles published between 2011-2013 and 2015-2018 to determine the impact of the checklist by Keil et al. (2014), on reporting practices. In their review of 150 ERP articles in five prestigious journals, equivalence tests showed that the impact of the checklist on methods reporting was statistically equivalent to zero. Based on this, the authors conclude that "*in short, the publication of the ERP guidelines article had no impact on reporting behavior in the 3 years after publication*" (Clayson et al., 2019, p. 4). Šoškić et al. provide similar evidence over a wider time range, comparing the 25 N400 studies in their sample published prior to 2000 (when the guidelines by Picton et al. (2000) came out) and the 25 studies published since 2015 (after the guidelines by Keil et al. (2014) came out). The results have shown that the amount of ambiguous and omitted information was similar in both groups of articles, further supporting the conclusions of Clayson et al. (2019). Altogether, it seems that guidelines and current checklists have not facilitated clear reporting in the published literature. We argue that the shortness and ambiguity inherent in written reports of EEG methods limits transparency, replicability, reproducibility, and metascience.

4. What are the alternatives to finding methodological details in journal articles?

One potential way for a replicator or meta-scientist to overcome the challenge of incomplete/ambiguous reporting is to request additional information from the corresponding author of a published study.

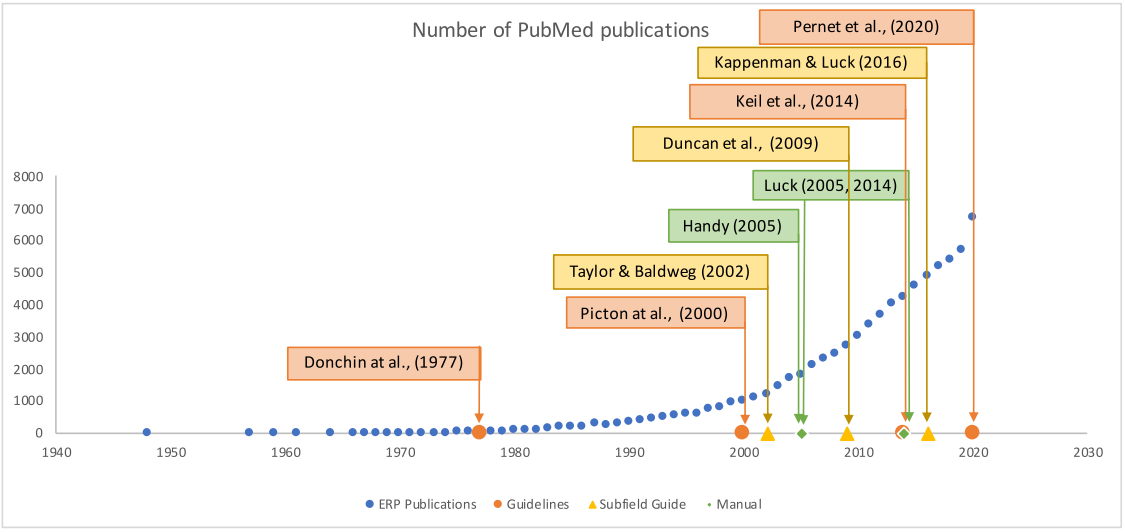

Fig. 1. Number of PubMed articles containing the terms 'ERP' and 'Event related potential(s)' per year (blue dots), along with the publication of notable general guidelines (red flags), subfield guidelines (yellow flags), and practice manuals (green flags).

However, recent investigations into the success of such requests have shown that the rate of successful attempts to obtain required information from corresponding authors in psychology ranged between 48–62% (Abernethy & Keel, 2016; Reid et al., 1982; Wimmer and Reid, 1982), and more than half of all requests did not receive any response —positive or negative— from authors in the fields of biomedical sciences (Manca et al., 2018; Teunis et al., 2015). This has led one group of meta-analysts to protest about the "non-corresponding" nature of corresponding authors (Manca et al., 2018).

An alternative approach has been to call for archiving research data along with code (or software session history), outside the body of the written journal article (e.g., in supplementary documents, or Open Access platforms). While such data is an invaluable resource for other researchers, some obstacles remain if the research objects have not been designed or stored with reproducibility in mind. For example, file or variable names may not be transparent; file or folder structures may obscure relationships between different parts of the data; code may be unusable on a different device, or written in a way that does not allow even the most experienced coders to understand its function (Community et al., 2019). Additional problems include compatibility between different versions of software, and the ability to access data types generated by some commercial software systems.

New movements that addresses these challenges include the development of the 'FAIR Guiding Principles for scientific data management and stewardship', which promote the idea that research data (including metadata about what has been done in the course of a scientific study), should be 'Findable, Accessible, Interoperable, and Reusable' (Wilkinson et al., 2016). Initiatives working to create FAIR data templates and formats in the field of neuroscience include the family of BIDS formats for brain imaging data (Gorgolewski et al., 2016; Holdgraf et al., 2019; Niso et al., 2018; Pernet et al., 2019) as well as user-friendly manuals about designing FAIR research objects for future reuse (Community et al., 2019). Even if all of this is done well, extracting methodological details from the raw research objects (i.e., data and code) still requires substantial effort for a human reader, suggesting additional value can be achieved through supplementary documents summarizing methodological decisions in human readable formats outside the raw data and raw code archives (Pernet et al., 2020).

If we plan to make progress in designing tools that help researchers supplement their written reports with FAIR supplementary materials outside the body of the journal article, we can ask two valuable questions about what tools are needed by neuroscientists working in the particular domain of EEG: First, what are the specific errors, ambiguities and omissions that are prevalent in the field of EEG methodology reporting; and second, how can we design reporting tools that effectively prevent those errors? While systematic reviews can provide evidence to address the first question, to address the second, we turn to two fields where reporting templates have been used to effectively reduce errors.

5. What can we learn from other fields?

Whenever humans are working in a highly complicated environment there is a chance for human errors (Gawande, 2009). The fields of aviation and surgery have both tackled the problem of human errors by developing tools for industry-wide adoption, and critically evaluating not only the effectiveness of different tools (i.e., how well they prevent errors), but also investigating what factors influence tool adherence (i.e., conditions that support or prevent adoption of a standardized tool). Lessons learned by these industries provide powerful insights into why current reporting guidelines may fail, and what can be done to make more effective progress.

In the domain of aviation, unforced pilot errors can result in dramatic outcomes, including loss of life. One approach to reducing unforced errors has been the use of pre-flight checklists that are mandated in most airspace jurisdictions. One important field-study of checklist adherence took into account both procedural and 'socio-technical' factors, and proposed design guidelines for the development of future tools to reduce human error (Degani and Wiener, 1993). Of their eleven guidelines, eight have direct relevance to the development of tools for reducing errors in the complex descriptions of EEG methodology. Table 1 introduces each of these points, along with a discussion of how it corresponds to current EEG reporting guidelines; how it could be adapted into design guidelines for the development of new tools to enhance methodological reporting; and the actions needed to achieve this outcome.

The core lessons that can be learned from the field of aviation are that reducing human error is possible when error prevention tools are integrated into the standard operating procedure of a team, and that the tools themselves can reduce error when they have been designed with the following features: 1. asking for specific details, not simply statements that something has been done, 2. integrating reporting into operational processes, 3. removing reporting from the 'production de-

Table 1
Guidelines for the development of tools designed to reduce human error. Excerpts from Degani & Wiener's guidelines for the development of aviation checklists to reduce error (1993, pp. 358) and their applicability to the development of tools for reporting EEG methodology.

Degani and Weiner (1993) guidelines for aviation	Current EEG Reporting Guidelines	ARTEM-IS Design Guideline for EEG Reporting Tools	Actions Required
D&G Item 1 Checklist responses should portray the desired status or the value of the item being considered, not just "checked" or "set".	Most current guidelines include statements like "report the filters clearly" which means that the response is a declaration about the completeness of the item, not the actual value of the item. Some current initiatives include reporting of values in the template body (Pernet et al., 2020), especially those that are designed to accommodate machine outputs from software systems (Gau et al., 2019; Gorgolewski et al., 2017). Guidelines about what one *should do* in EEG methodology are not always clearly differentiated from guidelines about how to report what one *has done* in EEG methodology.	**ARTEM-IS Item 1** Tools designed to support accurate reporting of EEG methodology should include fields where the item that needs to be reported is documented inside the tool itself - for example, in a data template, with the level of specificity required for direct replication, as informed by evidence of errors and ambiguities in the literature, such as are observed in systematic reviews. New tools designed to facilitate accurate reporting should not constrain what methods are possible for a researcher to perform, nor what order procedures must be performed in. Rather the template should provide a way of documenting what has actually been done in a clear format that is both *intuitive to fill* and *foolproof by design*.	New tools should be developed with item-level specificity for reporting the values of required items, using literature reviews to identify likely sources of error. New tools should be able to support flexible entry fields, and dynamic organization to allow for documentation of novel methodological procedures, settings or sequences.
D&G Item 3 A long checklist should be subdivided into smaller task checklists or chunks that can be associated with systems and functions within the cockpit.	Current guidelines are generally broken into sections associated with readily identifiable stages in EEG data acquisition, processing and analysis. Some ambiguities, errors and omissions are multi-stage decisions that are condensed into a single reporting item, rather than being broken down into their component steps. This is particularly common for descriptions of human decisions that may not be reflected in software logs or code (e.g., how an analysis window was chosen).	**ARTEM-IS Item 2** Tools designed to support accurate reporting of EEG methodology should be subdivided into smaller task checklists associated with readily identifiable stages in EEG data acquisition, processing and analysis. Error-prone reporting items should be further subdivided into smaller individual steps that are described more clearly, to ensure no steps are missing from the record, and that each step has the best chance of being reported accurately.	As in existing reporting guidelines, new tools should be structured with subdivisions of the reporting items into procedural stages. Literature reviews can be used to identify areas most in need of restructuring into smaller, clearer subdivisions.
D&G Item 4 Sequencing of checklist items should follow the geographical organization of the items in the cockpit and be performed in a logical flow.	The order of items in current guidelines generally follows the sequence of steps involved in EEG data acquisition, processing and analysis.	**ARTEM-IS Item 3** Tools designed to support accurate reporting of EEG methodology should order their items following a reasonable sequence of steps in EEG data acquisition, processing and analysis. When steps are subdivided into smaller-scale data items, a branching structure can be used to discretize the process of complex decisions into their component parts for greater clarity, and these elements should be sequenced so data items in one section constrain subsequent data items within the branch - as a form of logical check.	As in existing reporting guidelines, new tools should be structured with an order of subdivisions that reflects the order of acquisition, processing and analysis, and within a section, an order of items that reflects an optimal order of operations, or the typical order in which methodological decisions would be made.
D&G Item 5 Checklist items should be sequenced in parallel with internal and external activities that require input from out-of-cockpit agents such as cabin crew, ground crew, fuelers, and gate agents. We note here that this guideline could conflict with D&G Item 4.	Current guidelines specify documentation of some external details such as equipment and software, but the responsibility for providing these details lies with the person reporting the details, not with the supplier, and many details are difficult to find in the documents or user interfaces shipped with equipment and software. A number of current initiatives are in progress to create common data structures across the neurosciences (Pernet et al., 2020), including machine readable data objects in a common format (Gorgolewski et al., 2017), in line with FAIR Principles (Wilkinson et al., 2016).	**ARTEM-IS Item 4** Tools designed to support accurate reporting of EEG methodology will be more effective if they can integrate with information from hardware/software suppliers through the use of common formats or prefilled template sections. In cases where one technical detail may be described in different ways by commonly used software systems (e.g., filter settings), the tools developed for methodological description should allow data entry in any common format, and facilitate correspondence between formats.	Where possible, suppliers of software and hardware should be invited to coordinate on shared metadata formats by shipping prefilled templates along with their products to provide better pathways to accurate reporting. Common software systems can be reviewed for how methodological settings are described in their user interfaces, and tools should be designed to facilitate correspondence between systems using common metadata structures.
D&G Item 6 The most critical items on the task checklist should be listed as close as possible to the beginning, in order to increase the likelihood of completing the item before interruptions may occur. We note that this guideline could conflict with D&G Items 4 and 5. In most cases when this occurs, this guideline (D&G Item 6) should take precedence.	In all current reporting Guidelines, the order of items follows a common order for operations to be performed (as in Item 3), suggesting that no items have precedence in terms of importance. However, according to Šoškić et al. (2021), some items are reported more frequently by authors of EEG studies, suggesting that they have priority in the minds of authors, editors and reviewers of journal articles. The writers of reporting guidelines do not suggest that some items are more critical than others. However, systematic reviews show that some methodological items do indeed have priority over others in current reporting traditions.	**ARTEM-IS Item 5** If some reporting items are deemed 'more critical' than others, it may be possible to have some mandatory and some optional reporting items. However, the decision about whether some reporting items are indeed more crucial than others will require agreement of the community at large.	Developers of new tools should identify mismatches between 'Practice Guidelines' and habits in the existing literature during tool design, and engage in consultation with the community to determine which reporting items (if any) should have priority in a standardized template. Future metascientific approaches may help to refine the model of which reporting items have the largest effect on study outcomes, and are therefore most critical to replicability.

(*continued on next page*)

Table 1 (*continued*)

Degani and Weiner (1993) guidelines for aviation	Current EEG Reporting Guidelines	ARTEM-IS Design Guideline for EEG Reporting Tools	Actions Required
D&G Item 8 The completion call of a task checklist should be written as the last item on the checklist, allowing all crew members to move mentally from the checklist to other activities with the assurance that the task checklist has been completed.	Current Guidelines are effectively 'completed' when a researcher decides that every statement has been fulfilled, or 'ticks off' every statement in a checklist. There is no standardized procedure for checking that the items have been completed in line with the recommendations in the published guidelines, nor for coordinating the end of the checking process with other milestones in the reporting pipeline (e.g., before, during or after revisions arising from peer-review). Rather the requirement for accurate reporting is constant but never procedurally 'complete'.	**ARTEM-IS Item 6** Tools designed to support accurate reporting of EEG methodology should incorporate internal validity checks that help a researcher to validate whether the items have been completed in line with published guidelines, with a particular focus on those errors observed in reviews of the existing literature. Three types of check are recommended: a check that all required methodological details *have been reported*; a check that all required methodological details *have been reported with valid values* (e.g., a filter cutoff reported in Hz); and a check that methodological details in one section of the template *are logically compatible* with details entered in another section of the template (e.g., if a practitioner reports using a pre-configured electrode cap, then the electrode placements should match the make and model of the cap). Different validity checks may be relevant at different stages in methods reporting (e.g., at the time of preregistration versus at the time of submission for publication).	New tools for documentation should take the form of a digital template that will allow automated checks of whether required fields have been completed. The fields should be designed to incorporate item-specific data entry formats to facilitate entry of correct values (e.g., numerical values only for some fields). It is recommended that automated validity checks generate *informative notifications* about the status of data entered in a particular field (e.g., "no errors detected"; "item 5 incompatible with item 7"; "item incomplete"). When a template is declared 'complete' by a practitioner, the template should include an internally generated completion report including potential conflicts in internal logic. To remain consistent with **ARTEM-IS Item 7** (recovery from error), it should be possible to create and save an interim template with incomplete or inconsistent data items for later clarification.
D&G Item 10 Checklists should be designed in such a way that their execution will not be tightly coupled with other tasks. Every effort should be made to provide buffers for recovery from failure and a way to "take up the slack" if checklist completion does not keep pace with the external and internal activities.) **D&G Item 11** Flight crews should be made aware that the checklist procedure is highly susceptible to production pressures. These pressures set the stage for errors by possibly encouraging substandard performance and may lead some crew members to relegate checklist procedures to a second level of importance or not use them at all.	The phrasing of current guidelines suggest that documentation of methodological details may be done just once, at the time of manuscript preparation. This conflates methodological documentation with the task of article preparation, and the production demands that go along with writing for publication (e.g., time pressure, page counts). In this pipeline, there is little room for 'recovery from failure', as any declaration that "guidelines have been followed" may be overturned later in the drafting or revision process. Reporting of EEG methodological details can be impacted by production pressures like page limits in traditional journal formats, as well as numerous rounds of drafting and revising among multiple author teams. An additional production pressure arises from the limitations of describing technical details in concise written descriptions. In fields outside EEG, protocols exist for a variety of data objects that describe methodological details outside the body of the journal article, including templates for reporting systematic review and meta-analysis (Lasserson et al., 2021; Page et al., 2021), and pre-registration (Sullivan et al., 2019; WhartonCredibilityLab, 2015), and field-wide adoption of these tools has been encouraged by prominent journals requiring these documents before considering an article for publication.	**ARTEM-IS Item 7** Tools designed to support accurate reporting of EEG methodology should allow completion of different parts of the checklist at different times, should facilitate continuous revision to improve accuracy, or recover from 'error'. For this reason, the task of methods documentation should be separate from the task of writing, drafting, proofing and revising the research report in journal article format. Separating the tasks of documenting technical details from journal article writeup also facilitates **ARTEM-IS Item 8**. **ARTEM-IS Item 8** To separate the task of methodological documentation from the production demands of article writing, new tools should allow full and accurate documentation outside the body of the journal article, as consistent with **ARTEM-IS Item 7**. As in other fields, such tools should be considered as supplementary documentation, archived alongside other deliverables arising from the research (e.g., a persistent archive for the project; a journal's supplementary documents). Tools should be designed to reduce the burden of separate reporting as much as possible by a) skipping/hiding reporting items that are not logically possible given the answers to earlier reporting items, and b) designed to facilitate reuse and sharing of common pathways/pipelines - in line with FAIR Principles.	New tools for documentation should take the format of a meta-data template that can be completed independent from research reports for journal publication, and should be designed to allow revision (e.g., creation of an incomplete template, or reports that may contain inconsistencies at an early stage in reporting), and allow revision up until the time of version-controlled archiving. New tools should provide data templates for creation of supplementary documentation in a standardized format. To address the challenge of whether accurate reporting is considered a priority, when community standards have been agreed, stakeholders in EEG research can be consulted about mandatory reporting obligations - for example, making reporting via agreed templates a requirement of publication and/or reporting to funding agencies.

mands' of activities like journal article writing, which have high value in academic promotion and tenure decisions; 4. allowing revision to 're-cover from failure' and 5. confirming that reporting is 'complete' via item level validity checks, and internal logic checks.

Aviation checklists were also the inspiration for the Surgical Safety Checklist developed by the World Health Organization. After conducting a series of reviews into surgical safety in 2004, the WHO developed the Surgical Safety Checklist (WHO, 2009), designed to reduce human errors in the high-complexity environment of operating theatres. The first evaluation was conducted in parallel by eight hospitals in eight different countries, and in the first 30 days of use, the checklist was shown to significantly reduce the overall number of human errors, and the surgical

mortality rate (Haynes et al, 2009). Ten years after the introduction of the Surgical Safety Checklist, it was shown that those teams who use the checklist most often have the largest reductions in post-operative complications and mortality (Klai, 2019). Follow-up studies have looked at some of the challenges of introducing the WHO Surgical Safety Checklists in different environments, including initial reactions from different stakeholders in the surgical procedure. Three main themes emerge that can limit adoption of checklists: organizational hierarchy; momentum following rollout; and operational barriers to adoption stemming from the perceived value of the tool given the time taken to complete it.

Several studies have now noted that hierarchy can be a barrier to effective checklist adoption. In this context, people who have less prestige in the operating theatre are often highly supportive of standardized tools that can help to streamline complex procedural sequences. By contrast, people with more prestige tend to have higher confidence in their procedural knowledge, and see less value in adding checklists to their work flow. In one early pilot of the checklist in the UK (Vats et al., 2010), nurses were generally supportive of the checklist, but more pushback was experienced from senior surgeons. Subsequently, checklist adoption was poor in the absence of support from senior clinicians. A second, large-scale study of checklist adoption across 18 centres in France (Fourcade et al., 2012) also identified organizational hierarchy as a barrier to checklist adoption. A survey about checklist use at a large hospital in Austria showed that even though all members of operating teams showed equivalent levels of objective knowledge about checklists, surgeons tended to overestimate their subjective knowledge relative to anesthetists and nurses (Sendlhofer, 2015). From the first of these studies, the authors' recommend providing training materials that highlight the advantages of tool use, along with cultivating local champions to support adoption of new tools (Vats et al., 2010).

Despite the fact that studies routinely show that checklists reduce postoperative complications and mortality (Haynes et al, 2009; Klai, 2019), some studies have shown a loss of momentum following initial rollout of a checklist. One study (Sendlhofer, 2015) revisited surgical teams some months after rollout, and found that checklist use dropped rapidly after early adoption if it was not actively promoted by the organization. The authors proposed that active monitoring through external 'audits' might help to improve compliance, alongside promotion by local 'champions'. However, the loss of momentum seemed to stem mainly from operational pressures, such as time pressures for the operating team.

From an operational perspective, skilled professionals may resist adopting a new checklist if they see it as a disruption to their well-honed procedures. This is particularly the case when a checklist is perceived as a 'tick-box' exercise that causes extra workload without adding value, for example, when there is duplication between checklists. (Fourcade et al., 2012). One recommendation to address operational opposition is to build community buy-in by involving practitioners in the development of local adaptations of the tool. Some propose that adaptations should be limited to essential changes, and should avoid oversimplification or modification 'for the sake of it' (Vats et al., 2010). A particularly effective form of localization is integration of the checklist with other technical systems in the operating theatre - in one study, integrating an 'Electronic Surgical Safety Checklist' into the existing digital monitoring systems almost doubled checklist use, and significantly reduced postoperative complications (Gitelis et al., 2017) - a powerful demonstration that reducing the operational barriers to checklists is the best pathway to successful adoption.

Clearly, errors, ambiguities and omissions in the reporting of a single EEG study are unlikely to have the same kind of life-and-death outcomes as in aviation and surgery. However, inaccurate and non-replicable research certainly slows down the general progress of neuroscience as a field, and may generate a variety of scientific, ethical and financial consequences downstream.

6. What do we need to do?

The takeaways from these studies have shown that checklists or templates that help us to perform our duties to the best possible level of precision do help to reduce errors. Unsupportive leadership can be a barrier to best practice in following guidelines that reduce errors, and people with the greatest expertise have a tendency to overestimate their knowledge and their procedural recall of fine grained steps or details they use in their daily practice. These are the people most likely to consider the addition of a checklist to be an unnecessary burden due to operational pressures (time pressure to produce journal publications given current incentive structures in academia). These pressures are a legitimate concern that can only be addressed if tools developed to enhance reporting accuracy make the job of reporting easier rather than harder — for example by targeting common errors, omission and ambiguities, integrating automated data validity checks, and including informative error detection in checklist design. Good training materials that explain the benefits of a detailed reporting checklist may help with adoption, along with best practice champions.

Working together as a community of end users will provide the best pathway to developing tools with the greatest utility. To launch our collaborative consultations, in October 2020 we announced the ARTEM-IS Statement (https://osf.io/mf97q/) at the LiveMEEG online conference (https://livemeeg2020.org) - as a call to action for the development of an Agreed Reporting Template for EEG Methodology, that can be developed into an International Standard. Members of the broad community of stakeholders in EEG are invited to become signatories to the statement, as an indication of their commitment to improve reporting standards through collective action to develop new tools that make reporting easier through application of systematic Design Guidelines. To date the ARTEM-IS Statement has 43 signatories.

7. The ARTEM-IS statement

Accurate reporting of scientific methodology is critical for the scientific record. Due to the complexity of electroencephalography, EEG research contains multiple degrees of freedom in different stages of data acquisition, processing and analysis. This allows large-scale flexibility, and approaches can vary greatly between different research contexts or projects. Technical reporting of steps in the methodology is also highly convoluted, and verbal descriptions can obscure procedural details. Over the past four decades, researchers have made advances in Guidelines outlining best-practice for reporting standards (Donchin et al., 1977; Keil et al., 2014; Picton et al., 2000). However, recent reviews have shown that reporting often contains ambiguities or omissions (Ke et al., 2021, in progress; Šoškić et al., 2021). Furthermore, inconsistencies in the way particular processes are described make it hard to compare different descriptions of the same method, or to group together research on the basis of methods used. Together these challenges mean that the current literature is unsuited to detailed direct replications or methodologically informed meta-analysis.

We believe that action is necessary to help researchers achieve better reporting standards by designing tools that facilitate detailed methodology documentation, thereby reducing the likelihood of errors and omissions in reporting. This outcome will be achieved more effectively with cooperation of stakeholders across the community who work with EEG products, projects and outputs, to ensure ease of use, clarity and specificity relevant across sectors of the community including hardware manufacturers, software developers, researchers, journal editors, and data archivists. A print-ready PDF of this statement, and a draft template for community consultation towards an ARTEM-IS template for ERP can be found here (https://osf.io/pvrn6/).

Signatories to this statement agree with the following:

1 A standardised reporting tool can improve transparency and clarity of reporting for EEG methodology. This helps strengthen the scientific record and supports future scientific progress including conceptual replications, direct replications and metascience.
2 A standardized template can make reporting of EEG methodology easier, and provide technical details that are less ambiguous, with lower likelihood of omissions or errors.
3 A digital reporting template can be used to archive methods details alongside original research documentation such as protocols, data and code, to document the full research cycle without exhausting limited space in traditional publication formats.
4 A standardised reporting tool will be beneficial to the community of stakeholders in EEG research, including frontline EEG researchers, EEG hardware and software designers, journal editors, advocates for open and replicable methods, meta-analysts, and data-archivists.
5 The community of stakeholders in EEG research can develop more powerful tools by working together. Examples include hardware manufacturers providing machine specifications as pre-filled fields in the reporting template; software developers producing template-ready summaries from code, and data archivists specifying the metadata formats of greatest future utility.

While several initiatives in neuroscience are currently working towards the goal of transparent, FAIR metadata in neuroscience (Gorgolewski et al., 2016; Holdgraf et al., 2019; Niso et al., 2018; Pernet et al., 2019), many of the current templates require written descriptions of complex methods, and as such, suffer from similar limitations to those found in the existing guidelines for written descriptions in journal articles: Many reporting formats allow vague or errorful information to be filled. By contrast, the ARTEM-IS approach prioritizes using ***evidence of errors*** to guide the design of a tool that is ***intuitive to complete*** and ***foolproof by design***. This kind of tool thereby helps a practitioner to complete the template accurately, with the correct information type in the expected location, and checks for whether information is valid and logically consistent. In the section that follows we give an example of how the ARTEM-IS Design Guidelines can be applied in tool design to prevent a particular type of error observed in the published literature. We will take as our target how a researcher reports their selection of dependent variables over particular channels and time windows, for analysis of one subdomain of EEG - event-related potentials (ERPs).

8. Applying the ARTEM-IS Design Guidelines to create an evidence-based template for ERP

In the 2014 publication guidelines for EEG and MEG, Keil (2014, p.20) includes a reporting checklist with yes-no statements about whether each of three items have been reported: 1. “Measurement procedures are described, including the measurement technique (e.g., mean amplitude), the time window and baseline period, sensor sites, etc”; 2. “For peak amplitude measures, the following is included: whether the peak was an absolute or local peak, whether visual inspection or automatic detection was used, and the number of trials contributing to the averages used for measurement”; 3. “An a priori rationale is given for the selection of time windows, electrode sites, etc.” The full text of the article contains further elaboration on how to complete these items in the body of the journal article, while the checklist acts as a supplement outside the body of the article. One limitation of this design is that it relies on a practitioner to check that they have completed the item correctly, and that the completion of the self-check does not guarantee transparency or accuracy. Another limitation is that many methodological decisions are covered in a single checklist step.

More recently, the COBIDAS MEEG Best Practices Guidelines (Pernet et al., 2018b) provides a table of recommendations, where they break down the same reporting obligation into two rows, consisting of five reporting items: “ROIs. 1. How were they determined, i.e. what was the mode of selection (e.g., a priori from literature or independent data)? 2. Report specific sensors/regions of interest, peaks, components, time and/or frequency, window, source; Summary Measures: 1. Report how these were obtained; 2. Justify how the selection of dependent variables is unbiased (especially how the temporal and spatial ROIs were chosen); 3. Describe how peaks, components, latencies were measured.” (p.48). As in the case of the Keil checklist, in order to complete the reporting items with accuracy, the full text of the article needs to be consulted as a kind of further elaboration. The elaboration is detailed, and helpful as it combines justifications for *why a particular detail is necessary* with *the required details to be included* in a report. However, the full text format relies on a reader to conduct their own self-check of whether these items have been completed to the expected level of detail, and each of the five items contains several methodological details which are not further broken down into discrete elements.

Two main features differentiate ARTEM-IS from these approaches to EEG methodology reporting: Using an evidence-based approach for error prevention, and applying the ARTEM-IS Design Guidelines for developing new tools. An example of this process of combining evidence of errors with the Design Guidelines follows.

EVIDENCE-BASED TARGETS FOR REDUCING ERROR. In Šoškić et al. (2021), we observed that contrary to current reporting guidelines, 36% of N400 articles failed to describe how they decided on a time window for analysis (i.e., a-priori, data-driven, etc.). A further 5% described a-priori decisions based on previous literature or experience, but gave no specific sources. 26% of the total did cite specific sources in the previous literature, but in more than half of those cases, the window(s) in the cited source did not match the window in the article precisely, and no further elaboration was given about the discrepancy. Similar ambiguities and omissions were observed for articles reporting on data-driven approaches. In order to reduce these specific errors, items outlined in existing guidelines were targeted for detailed attention in a new draft tool.

DESIGN PRINCIPLES FOR TEMPLATE DESIGN. The ARTEM-IS Design principles help to create a template that reduces these specific errors, as each area of reporting is broken down into multiple smaller data-items (**ARTEM-IS Item 2**), elaborated using a branching decision structure (**ARTEM-IS Item 3**), where actual values to be reported are informed by errors and omissions in the literature (**ARTEM-IS Item 1**). To reduce production demands on the form filler, the majority of data reporting items are skipped/hidden from view, depending on earlier choices (**ARTEM-IS Item 9**), resulting in a concise decision tree (See Fig. 2). Two ongoing goals are to build a user interface that includes further internal validity checks (**ARTEM-IS Item 6**), and pre-filled template sections that populate common standards or settings from one part of the template to another (**ARTEM-IS Item 4**).

In this example, detailed evidence about the specific types of errors in 132 articles within this particular field of MEEG helped us to design question-types and data checks that preclude precisely those errors and ambiguities. The branching question structure shown in Fig. 2 prompts a practitioner to include the level of detail expected in various professional guidelines (Keil et al., 2014; Pernet et al., 2018b), by breaking down the reporting items into over 20 discrete decision points or data entry fields. This decision tree incorporates validity checks that can help to ensure that the information entered is logically consistent (such as requiring a practitioner who reports decisions based to previous literature to cite the literature in question), along with prompts for further elaboration in cases where necessary details are frequently omitted (such as elaborating on how discrepancies in the literature were handled). This combination of evidence base for errors and application of Design Guidelines aims to make the template ***intuitive to complete*** and ***foolproof-by-design***.

The detailed elaboration provided by the ARTEM-IS template for ERP may appear to be more time consuming than a template with fewer re-

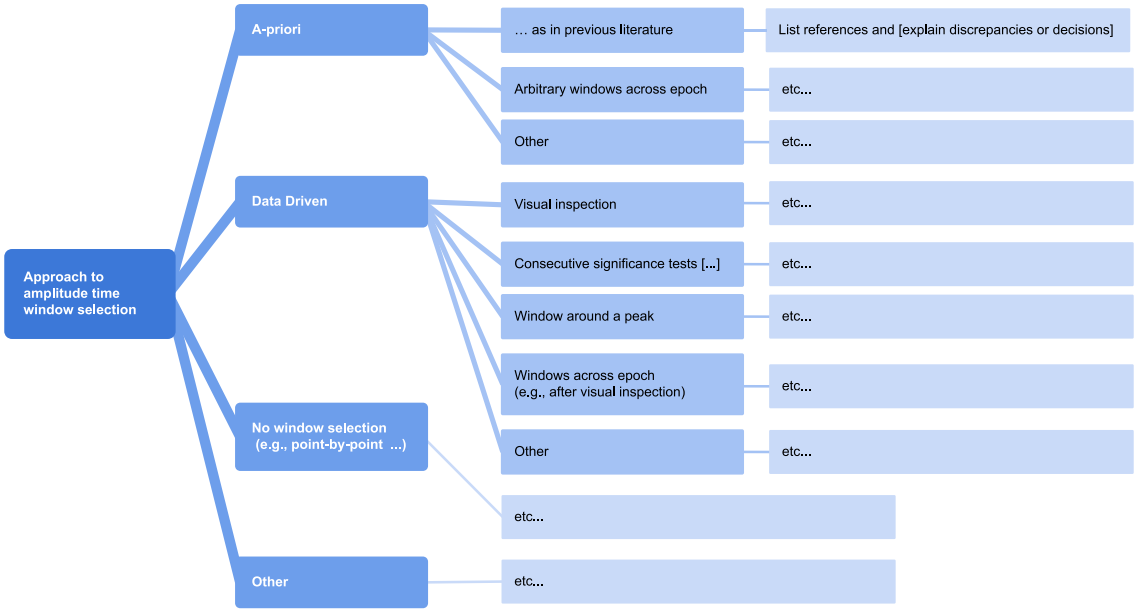

Fig. 2. Schematic of a decision tree showing how reporting the "Approach to time window selection" is broken down into multiple, smaller scale reporting items to reduce the likelihood of omission or error. A person completing the form only sees those items relevant to earlier decisions, helping to streamline the process of reporting, and reducing logically impossible combinations of reporting items. The branching structure mirrors the nature of the Garden of Forking Paths in methodological decisions (Gelman and Loken, 2013).

sponse fields. However, we believe that discretizing individual reporting items into a sequence of simple decisions may actually save a practitioner time, since they are prompted to include each detail in sequence, and internal logic checks reduce the need to consult published guidelines for compliance. Furthermore, although the initial overhead may take some time, any research project or team that uses a common pathway across several studies can reuse or adapt a pre-filled template describing common operations, and update only those elements that deviate from previous pipelines. Not only does this approach save time in documentation, but also reduces the problem of whether a methods section is 'self plagiarised' between articles, as methodological documentation can be cited as a discrete data object, and details required in the text of a written methods section can be extracted directly from the template. Future extensions of such a template may include the generation of automated prose descriptions from template values, further reducing the burden of methodological reporting. In addition, for researchers who plan to pre-register their work, or coordinate research activities across disparate research sites, the item-level specificity of fields within an ARTEM-IS template may also simplify the pre-registration process, by providing validity checks for items prior to data collection. This approach to item-level specificity can help to reduce researcher degrees of freedom by specifying a researcher's proposed path through the garden (Šoškić et al., 2019). We hope to see the ARTEM-IS approach complementing other endeavours towards pre-registration in the field of EEG (e.g., Paul et al., 2021).

In accordance with this initiative, our first target is to develop a template suitable for reporting ERPs, based on the detailed evidence base accumulated through systematic reviews about error types in reporting two ERP components (Ke et al., 2021 in progress; Šoškić et al., 2021). Although the N400 and MMN are niche components within the field of ERP research, the detailed data structure being developed as part of the ARTEM-IS template may be adapted further for other subfields of EEG (hence the 'E' in ARTEM-IS stands for EEG, broadly construed). Through this process, we have developed guiding principles for the design of error-reducing tools that are broadly compatible with other community efforts to develop common metadata templates for neuroscience and beyond.

9. Community consultation and coordination

At the time of announcing the ARTEM-IS Statement, we shared a draft of an ARTEM-IS template for ERP research to serve as the starting point for development of agreed standards. The draft template contains nine sections: (1) title, (2) hardware (equipment), (3) data acquisition, (4) pre-processing, (5) experimental design and sampling, (6) measurement, (7) channel (electrode) selection, (8) results visualization, and (9) other, with an ongoing discussion regarding a potential tenth section for statistical analysis. Each section includes a set of specific items which allow describing the methodological decisions in detail by filling in appropriate numerical and categorical values, in a branching decision structure. The original draft version of the items included in the template is archived on the ARTEM-IS Open Science Framework project (link to the template: https://osf.io/w4nt6/). The initial ERP template acts as a pilot for the ARTEM-IS template and design process, based on which templates for other subfields may follow.

This bottom-up approach to error reduction in a narrow field of EEG has a natural synergy with other initiatives currently underway in the reporting of neuroscience methodology - many of which take their structure from shared macro-level concerns that apply across the neurosciences and other related disciplines. For example, the Organization for Human Brain Mapping (https://www.humanbrainmapping.org) is notable for the global network contributing to the common goal of improving neuroscience methodology, and their commitment to a common toolkit for reporting, covering a variety of neuroimaging fields (Nichols et al., 2017; Pernet et al., 2020; Pernet et al., 2018b), coordinating with machine-readable data-types like BIDS (Appelhoff et al., 2019; Delorme et al., 2021; Gorgolewski et al., 2017; Holdgraf et al., 2019; Niso et al., 2018), and working towards common lexicons that can help to reduce ambiguities in the description and interpretation of methods (Helmer et al., 2018; Pernet et al., 2020). The COBIDAS MEEG community has made progress turning reporting guidelines into a human fillable, machine readable data template with an app interface called eCOBIDAS (Gau et al., 2019). We see the natural interface between these two initiatives as the intersection between the shared lexicon and common reporting framework provided by the COBIDAS fam-

ily of reporting tools, and the detailed, branching logic of the ARTEM-IS templates that are targeted to reduce evidence-based errors in specific subfields. We look forward to further integration of these approaches as we share common formats, and work collectively towards the goal of making FAIR reporting achievable in neuroscience.

Since the presentation of the ARTEM-IS Statement at the LiveMEEG (Šoškić et al., 2020), further development has begun through collaboration of the broader community of stakeholders: The ARTEM-IS team has expanded and formalized into an International Neuroinformatics Coordinating Facility (INCF) Working Group (https://www.incf.org/sig/incf-working-group-artem). The ARTEM-IS INFC Working Group holds frequent meetings for all members to attend, with the stated goal to "create human-user friendly, machine-readable templates for documenting the methodological details of an EEG study or pipeline, and to build these templates into usable tools". The Working Group is open for new members to join. The Working Group has also partnered with eCOBIDAS (Gau et al., 2019) to transform ARTEM-IS templates into a clickable web-app which will allow easier completion of the template using an onscreen interface. After the first ARTEM-IS OHBM BrainHack (Styles et al., 2021) several reporting sections now have working prototypes within the App interface. Further steps of the Working Group will include developing app functionality to allow data export (both machine-readable .json outputs, and human-readable reports), and import (e.g., default values provided by a vendor of equipment or software; a research project's standard operating procedure). A further goal of the Working Group, is to develop guidance to support new users of the app, such as app-integrated 'help' functions drawing from community developed tools and resources (e.g., links to shared lexicon of terms); and training materials for researchers new to the field of EEG. With these goals in mind, the ARTEM-IS INCF Working Group aims to create tools that increase transparency and accuracy in the scientific record by implementing human-centred design guidelines based on the experience of other high complexity fields, to create tools that are ***intuitive to fill***, and make the job of reporting ***foolproof-by-design***.

10. Conclusion

EEG research has increased over the years, alongside calls for greater accuracy in reporting, but the publication of more and more detailed guidelines does not seem to provide a solution for the errors, omissions and ambiguities found in a substantial proportion of the published literature. Rather, the problem seems to stem from the production demands of performing methods documentation at the same time as writing scientific reports, in a format that is not suited to high accuracy, and does not provide a pathway for error detection. In the complex fields of neuroscience in general and EEG in particular, lack of accuracy in methodology reporting means that it is not possible to determine what path has been taken through the garden of forking paths, and this has consequences for transparency in the scientific record, as well as replication science, and metascience in the field. In other high complexity fields, the best solutions seem to come from creating tools that help practitioners reduce errors by providing a systematic framework —or template— with the following features: Item-level specificity (e.g., reporting actual values, rather than simply that a step has been performed correctly); Coordinating the process of reporting with other procedures that are performed, and other procedural systems; Separating reporting from the production demands of high stakes activities (i.e., separating detailed methodological reports from the body of journal article write-ups); and creating tools that make the job easier to do well. These procedures have higher take-up when the community of end users is involved in the development of the tool, and the tool is championed by some high-profile users within the community. We believe that these goals can be met most concretely when evidence from the existing literature is used to target specific errors that occur frequently in specific subdomains of neuroscience (here, the event-related potentials branch of EEG), but the framework developed in one area of specialization may have applicability for the neurosciences more broadly.

To meet this need within the context of EEG methodology we have proposed a series of principles that should be incorporated into the development of new tools for accurate reporting. The ARTEM-IS Statement outlines five principles: A standardised template can (1) **improve clarity** of reporting in the scientific record by providing a tool that helps to (2) **improve the accuracy** of individual item reporting. In so doing, such a template can (3) **enhance documentation** without using up limited reporting space, thereby bringing (4) **broad benefits** to the community of EEG stakeholders, but the development of such a tool will require (5) **community effort**. By way of elaboration, we drew on the high-complexity fields of surgery and aviation in the creation of the ARTEM-IS Design Guidelines for creating tools that target specific, evidence-based errors in the field of EEG. The ARTEM-IS template is designed to be co-developed through broad consultation with the community of stakeholders in EEG. Signatories to the ARTEM-IS Statement agree to be part of this collective endeavour, and members of the EEG community are invited to join the ARTEM-IS INCF Working Group to take part in these ongoing efforts.

Data & code availability

N/A.

Acknowledgements

This research was supported by the following funding sources: Singapore's National Research Foundation under the Science of Learning call (NRF2016-SOL002-011), Nanyang Technological University, via the Centre for Research and Development in Learning, (JHU IO 90071537) and NAP-SUG to SJS (M4081215), and grant OI179033 by Ministry of Education, Science and Technological Development of the Republic of Serbia.

The authors would like to thank Maximillien Chaumon and the organizing committee of LiveMEEG for the opportunity to present ARTEM-IS in an invited symposium at the LiveMEEG, which immensely helped kick off this project, as well as our discussants in the session, Emily Kappenman, Bertille Somon and Giorgio Ganis.

Authorship contributions

SJS, VK, HK and AS created the draft ARTEM-IS template and wrote the ARTEM-IS Statement; SJS and AS wrote the paper with contributions from VK and HK.

References

Abernethy, A.M., Keel, A.L., 2016. How researchers respond to replication requests revisited. J. Advert. 45 (1), 13–18. doi:10.1080/00913367.2015.1079751.

Appelhoff, S., Sanderson, M., Brooks, T., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A., Larson, E., Gramfort, A., Jas, M., 2019. MNE-BIDS: organizing electrophysiological data into the BIDS format and facilitating their analysis. J. Open Source Softw. 4 (44), 1896. doi:10.21105/joss.01896.

Bishop, D. (2017). Reproducible practices are the future for early career researchers. *BishopBlog.* http://deevybee.blogspot.com/2017/05/reproducible-practices-are-future-for.html.

Boudewyn, M.A., Luck, S.J., Farrens, J.L., Kappenman, ES., 2018. How many trials does it take to get a significant ERP effect? It depends. Psychophysiology 55 (6), e13049. doi:10.1111/psyp.13049.

Clayson, P.E., Carbine, K.A., Baldwin, S.A., Larson, M.J., 2019. Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. Psychophysiology 56 (11). doi:10.1111/psyp.13437.

Community, T.T.W., Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O'Reilly, M., & Whitaker, K. (2019). *The turing way: a handbook for reproducible data ScĀ.* Zenodo. 10.5281/ZENODO.3233986

Degani, A., Wiener, E.L., 1993. Cockpit checklists: concepts, design, and use. Hum. Factors J. Hum. Factors Ergon. Soc. 35 (2), 345–359. doi:10.1177/001872089303500209.

DeLong, K.A., Urbach, & Kutas, M. (2017). *Concerns with Nieuwland et al. (2017).* University of California. http://kutaslab.ucsd.edu/pdfs/FinalDUK17Comment9LabStudy.pdf

DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat. Neurosci. 8 (8), 1117–1121. doi:10.1038/nn1504.

Delorme, A., Truong, D., Martinez-Cancino, R., Pernet, C., Sivagnanam, S., Yoshimoto, K., Poldrack, R., Majumdar, A., Makeig, S., 2021. Tools for importing and evaluating BIDS-EEG formatted data. In: Proceeding of the 10th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 210–213. doi:10.1109/NER49283.2021.9441399.

Donchin, E., Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., Hillyard, S. A., & Sutton, S. (1977). Publication criteria for studies of evoked potentials (EP) in man: Methodology and publication criteria. In Progress in clinical neurophysiology: Vol. 1. Attention, voluntary contraction and event-related cerebral potentials (pp. 1–11). Karger.

Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Polich, J., Reinvang, I., Van Petten, C., 2009. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin. Neurophysiol. 120 (11), 1883–1908. doi:10.1016/j.clinph.2009.07.045.

Gau, R., Gould van Praag, C., van Mourik, T., Wiebels, K., Adolfi, F., Scarpazza, C., Ruotsalainen, I., Tepper, A., Sjoerds, Z., Simon, J., Klapwijk, E., Hortensius, R., Bartlett, J., & Moreau, D. (2019). COBIDAS checklist. 10.17605/OSF.IO/ANVQY

Gawande, A., 2009. The Checklist Manifesto: How to get Things Right, 1st ed. Metropolitan Books.

Gelman, A., & Loken, E. (2013). The garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Department of statistics, Columbia University. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gorgolewski, K.J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M.M., Churchill, N.W., Cohen, A.L., Craddock, R.C., Devenyi, G.A., Eklund, A., Esteban, O., Flandin, G., Ghosh, S.S., Guntupalli, J.S., Jenkinson, M., Keshavan, A., Kiar, G., Liem, F., … Poldrack, R.A., 2017. BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. PLOS Comput. Biol. 13 (3), e1005209. doi:10.1371/journal.pcbi.1005209.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., … Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 160044. doi:10.1038/sdata.2016.44.

Handy, T.C. (Ed.), 2005. Event-Related Potentials: A Methods Handbook. MIT Press.

Helmer, K., Keator, D., Auer, T., Ghosh, S., Maumet, C., Nichols, T.E., Smruti, P., Poline, J.B., 2018. Development of an ontology for the INCF neuroimaging data model (NIDM). Proceeding of the OHBM-24th Annual Meeting of the Organization for Human Brain Mapping.

Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B.L., Gorgolewski, K.J., Groen, I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C.J., Jas, M., Knight, R., Lachaux, J.P., … Hermes, D., 2019. IEEG-BIDS, extending the brain imaging data structure specification to human intracranial electrophysiology. Sci. Data 6 (1), 102. doi:10.1038/s41597-019-0105-7.

Ioannidis, J.P.A., Munafò, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P., 2014. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cognit. Sci. 18 (5), 235–241. doi:10.1016/j.tics.2014.02.010.

Kappenman, E.S., Luck, S.J., 2010. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology doi:10.1111/j.1469-8986.2010.01009.x.

Kappenman, E.S., Luck, S.J., 2016. Best practices for event-related potential research in clinical populations. Biol. Psychiatry Cognit. Neurosci. Neuroimaging 1 (2), 110–115. doi:10.1016/j.bpsc.2015.11.007.

Ke, H., Ković, V., Šoškić, A., & Styles, S.J. (2021) (in progress). *Systematic review of MMN for categorical perception of speech in children (preregistration)*. https://osf.io/p9jy3/.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2014. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography: guidelines for EEG and MEG. Psychophysiology 51 (1), 1–21. doi:10.1111/psyp.12147.

Lasserson, T., Churchill, R., Chandler, J., Tovey, D., Thomas, J., Flemyng, E., Higgins, J, 2021. Standards for the REPORTING of PROTOCOLS of new cochrane intervention reviews (PR1-PR44). In: Higgins, J., Lasserson, T., Chandler, J., Tovey, D., Flemyng, F., Churchill, R. (Eds.), Methodological Expectations of Cochrane Intervention Reviews. Cochrane https://community.cochrane.org/mecir-manual/standards-reporting-protocols-new-cochrane-intervention-reviews-pr1-pr44.

Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique. MIT Press.

Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, 2nd ed. The MIT Press.

Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54 (1), 146–157. doi:10.1111/psyp.12639.

Manca, A., Cugusi, L., Dvir, Z., Deriu, F., 2018. Non-corresponding authors in the era of meta-analyses. J. Clin. Epidemiol. 98, 159–161. doi:10.1016/j.jclinepi.2018.01.009.

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S.M., Hairston, W.D., Ferris, D.P., König, P., 2017. Systems, subjects, sessions: to what extent do these factors influence EEG data?. Front. Hum. Neurosci., 11 doi:10.3389/fnhum.2017.00150.

Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T.T, 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20 (3), 299–303. doi:10.1038/nn.4500.

Nieuwland, M.S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D.J., Rousselet, G.A., Ferguson, H.J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E.M., … Huettig, F., 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. ELife 7, e33468. doi:10.7554/eLife.33468.

Niso, G., Gorgolewski, K.J., Bock, E., Brooks, T.L., Flandin, G., Gramfort, A., Henson, R.N., Jas, M., Litvak, V., Moreau, J.T., 2018. MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. Sci. Data 5 (1), 1–5.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., … Yarkoni, T., 2015. Promoting an open research culture. Science 348 (6242), 1422–1425. doi:10.1126/science.aab2374.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., … Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ n71. doi:10.1136/bmj.n71.

Paul, M., Govaart, G.H., Schettino, A., 2021. Making ERP research more transparent: guidelines for preregistration. Int. J. Psychophysiol. 164, 52–63. doi:10.1016/j.ijpsycho.2021.02.016.

Pavlov, Y.G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C.S.Y., Beste, C., Bland, A.R., Bradford, D.E., Bublatzky, F., Busch, N.A., Clayson, P.E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Giorgio, G., He, X., Hinojosa, J.A., … Mushtaq, F., 2021. #EEGManyLabs: investigating the replicability of influential EEG experiments. Cortex doi:10.1016/j.cortex.2021.03.013.

Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R., 2019. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Sci. Data 6 (1), 103. doi:10.1038/s41597-019-0104-8.

Pernet, C.R., Garrido, M.I., Gramfort, A., Maurits, N., Michel, C.M., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., Puce, A., 2020. Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. Nat. Neurosci. 23 (12), 1473–1483. doi:10.1038/s41593-020-00709-0.

Pernet, D.C., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., Puce, A., 2018a. Best practices in data analysis and sharing in neuroimaging using MEEG. OSF Preprints doi:10.31219/osf.io/a8dhx.

Pernet, D.C., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., Puce, A., 2018b. Best practices in data analysis and sharing in neuroimaging using MEEG. OSF Preprints doi:10.31219/osf.io/a8dhx.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology 37 (2), 127–152. doi:10.1111/1469-8986.3720127.

Reid, L.N., Rotfeld, H.J., Wimmer, R.D., 1982. How Researchers Respond to Replication Requests. J. Consum. Res. 9 (2), 216. doi:10.1086/208916.

Sandre, A., Banica, I., Riesel, A., Flake, J., Klawohn, J., Weinberg, A., 2020. Comparing the effects of different methodological decisions on the error-related negativity and its association with behaviour and gender. Int. J. Psychophysiol. 156, 18–39. doi:10.1016/j.ijpsycho.2020.06.016.

Šoškić, A., Jovanović, V., Styles, S.J., Kappenman, E.S., Ković, V., 2021. How to do better N400 studies: reproducibility, consistency and adherence to research standards in the existing literature. Neuropsychol. Rev. doi:10.1007/s11065-021-09513-4.

Šoškić, A., Kappenman, E.S., Styles, S.J., & Ković, V. (2019). *Garden of forking paths in ERP research – effects of varying processing and analysis steps in an N400 experiment*. https://osf.io/6nqxy/

Šoškić, A., Ković, V., Savić, A., Kappenman, E., Ke, H., Styles, S.J., 2020. Towards an Agreed Reporting Template for ERP Methodology: A Draft Proposal and Roundtable Discussion October 8. LiveMEEG, Virtual doi:10.17605/OSF.IO/PDX6Y.

Styles, S.J., Šoškić, A., Ković, V., Ke, H., Gau, R., Pavlov, Y., Woon, F.T., Yang, Y.F., Fischer, N., Niso, G., Sareen, E., Algermissen, J., Sundaray, S., Tan, J., Yap, D., Ganis, G., & Shaw, A. (2021). *ARTEM-IS Web App OHBM Brainhack 2021*. 10.17605/OSF.IO/ARD3W

Sullivan, I., DeHaven, A., Mellor, D., 2019. Open and reproducible research on open science framework. Curr. Protoc. Essent. Lab. Techn. 18 (1). doi:10.1002/cpet.32.

Tanner, D., Morgan-Short, K., Luck, S.J., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition: High-pass filtering and artifactual ERP effects. Psychophysiology 52 (8), 997–1009. doi:10.1111/psyp.12437.

Taylor, M.J., Baldeweg, T., 2002. Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. Dev. Sci. 5 (3), 318–334. doi:10.1111/1467-7687.00372.

Teunis, T., Nota, S.P.F.T., Schwab, J.H., 2015. Do corresponding authors take responsibility for their work? A covert survey. Clin. Orthop. Relat. Res. 473 (2), 729–735. doi:10.1007/s11999-014-3868-3.

WhartonCredibilityLab. (2015). *AsPredicted*. University of Pennsylvania. https://aspredicted.org/

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., … Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (1), 160018. doi:10.1038/sdata.2016.18.

Wimmer, R.D., Reid, L.N., 1982. Researchers' response to replication requests. Journal. Q. 59 (2), 317–319. doi:10.1177/107769908205900223.