PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Computer-assisted analysis of routine electroencephalogram to identify hidden biomarkers of epilepsy: protocol for a systematic review
AUTHORS	Lemoine, Émile; Neves Briard, Joel; Rioux, Bastien; Podbielski, Renata; Nauche, Bénédicte; Toffa, Denahin; Keezer, Mark; Lesage, Frederic; Nguyen, Dang; Bou Assi, Elie

VERSION 1 – REVIEW

REVIEWER	Cynthia Sharpe Starship Children's Health, Neurology
REVIEW RETURNED	21-Aug-2022

GENERAL COMMENTS	Dear Authors, I very much look forward to reading your research when complete. This will be a very useful study. If possible it would be of particular interest to highlight the index tests which diagnose epilepsy in cases where there are no inter ictal spikes on EEG. Sub-analyses or specific report if possible of index test performance in populations where interictal spikes are known to be particularly non-specific (for example in cerebral palsy), would also be of interest. Minor typo correction suggestion: "Epilepsy affects over 65 million people worldwide".

REVIEWER	Camilo Garcia Cleveland Clinic Florida, Neurology
REVIEW RETURNED	25-Aug-2022

GENERAL COMMENTS	The study entitled "Computer-assisted analysis of routine EEG to identify hidden biomarkers of epilepsy: protocol for a systematic review" by Lemon E, et all. is a fantastic idea to assess quantitative methods available to evaluate ambulatory EEG. This will provide us with new tools, ideas and clear accuracy of current methods available in clinical practice. This assessment will be the pivotal for new methods to assess interictal EEG. It will provide us with the real limitations and future possible solutions. I think that the methods are appropriate for this study, the authors will divide the data according to different features including type of epilepsy, lesional vs non lesional, adults vs children, etc. However, I do not see a clear division between temporal vs extra temporal epilepsy since features and biomarkers may be different and sensitivity of the current computer-methods may be higher in either population. Finally, the study does not suggest a comparison between human vs “machine” accuracy or discrepancy. The study entitled "Computer-assisted analysis of routine EEG to identify hidden biomarkers of epilepsy: protocol for a systematic review" by Lemon E, et all. is a fantastic

	idea to assess quantitative methods available to evaluate ambulatory EEG. This will provide us with new tools, ideas and clear accuracy of current methods available in clinical practice. This assessment will be the pivotal for new methods to assess interictal EEG. It will provide us with the real limitations and future possible solutions. I think that the methods are appropriate for this study, the authors will divide the data according to different features including type of epilepsy, lesional vs non lesional, adults vs children, etc. However, I do not see a clear division between temporal vs extra temporal epilepsy since features and biomarkers may be different and sensitivity of the current computer-methods may be higher in either population. Finally, the study does not suggest a comparison between human vs "machine" accuracy or discrepancy. The main goal of artificial intelligence in medicine is to support our diagnosis and help us to make decisions based on accuracy. Therefore, those methods need to be compared with clinical judgment, which is the current standard.

REVIEWER	Farah Mohammad King Saud University
REVIEW RETURNED	19-Nov-2022

GENERAL COMMENTS	The article titled "Computer-assisted analysis of routine EEG to identify hidden biomarkers of epilepsy: protocol for a systematic review" is a systematic review or report. The publication needs a major revision. Clearly define the research gap and highlight the outcomes. All the findings need to address and explained in well-formatted figures, tables, and graphs.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Dear authors, I very much look forward to reading your research when complete. This will be a very useful study.

Comment 1: *If possible, it would be of particular interest to highlight the index tests which diagnose epilepsy in cases where there are no interictal spikes on EEG.*

Authors' response: We agree that the subgroup of EEGs with no interictal activity is of particular interest for this review. We added this subgroup analysis in the **Methods: Data synthesis** and **Methods: Analyses** section:

> "We will report the pooled proportion of patients with focal vs. generalized epilepsy, adult vs. children, structural vs. non-structural epilepsy, IEDs on EEG**,** and with specific epilepsy syndromes." (p. 11, l. 212)
> "We will perform subgroup analysis for the following variables: epilepsy type (focal, generalized), epilepsy etiology (structural vs. non-structural), presence of IEDs, age groups (children (< 18 y.o.), adults (≥ 18 y.o.)), epilepsy syndromes, extracted marker, and dataset used." (p. 12, l. 240)

Comment 2: *Sub-analyses or specific report, if possible, of index test performance in populations where interictal spikes are known to be particularly non-specific (for example in cerebral palsy), would also be of interest.*

Authors' response: We agree with the suggestion and made the change to the manuscript:

> "If possible, we will also perform a subgroup analysis for populations with a higher prevalence of IEDs without epilepsy (cerebral palsy, autism spectrum disorder, attention deficit disorder)." (p. 12, ll. 241–243)

Comment 3: *Minor typo correction suggestion: "Epilepsy affects over 65 million people worldwide".*

Authors' response: We have corrected the text (p. 4, l. 42).

Reviewer 2

The study entitled "Computer-assisted analysis of routine EEG to identify hidden biomarkers of epilepsy: protocol for a systematic review" by Lemoine E, et al. is a fantastic idea to assess quantitative methods available to evaluate ambulatory EEG. This will provide us with new tools, ideas, and clear accuracy of current methods available in clinical practice. This assessment will be the pivotal for new methods to assess interictal EEG. It will provide us with the real limitations and future possible solutions.

Comment 1: *I think that the methods are appropriate for this study, the authors will divide the data according to different features including type of epilepsy, lesional vs non lesional, adults vs children, etc. However, I do not see a clear division between temporal vs extra temporal epilepsy since features and biomarkers may be different and sensitivity of the current computer-methods may be higher in either population.*

Authors' response: We agree with the reviewer's suggestion regarding an additional subgroup analysis in patients with focal epilepsies, separating patients with temporal lobe epilepsy and extra-temporal epilepsy. We added the suggestion to the text:

> **"**We will also perform a subgroup analysis […] for extra-temporal vs. temporal focal epilepsy**."** (p. 12, ll. 237–239)

Comment 2: *Finally, the study does not suggest a comparison between human vs "machine" accuracy or discrepancy. The main goal of artificial intelligence in medicine is to support our diagnosis and help us to make decisions based on accuracy. Therefore, those methods need to be compared with clinical judgment, which is the current standard.*

Authors' response: With his comment, the reviewer raises two important questions: What are the clinicians' performances in predicting seizure risk? And how could computational analysis of EEG assist their clinical judgement?

For the first question, we are not aware of any study evaluating the predictive accuracy of clinical judgement for seizure recurrence risk. This would require a prospective study with long-term follow-up in which clinicians are asked to estimate seizure recurrence risk for a given patient. While this question is interesting, it is not in the scope of the proposed systematic review.

For the second question, the first step to introduce biomarkers into the clinical workflow is to understand their expected performances on real-world data. This is the objective of the systematic review. Following the reviewer's comment, we commented on this in the **Discussion**:

> "Translation of technology to clinical practice requires strong evidence based on high quality research. This review is important because it will establish the potential of automatic analysis of EEG as a diagnostic tool for epilepsy, and if evidence to support its use is lacking, it will identify the pitfalls that need to be overcome in future research. Also, by systematically describing current practices that are used by research groups, it will serve as a reference for new researchers in the field. Upon completion of this review, we will have a better understanding of the potential ways that automated analysis of EEG could be

integrated into the clinical workflow; this information will be valuable to anyone designing clinical studies on clinical-decision support systems for epilepsy." (p. 13, ll. 278–285)

In keeping with the idea of comparing "machine" with "human" performances, we changed the **Introduction** and **Discussion** to more explicitly reflect the fact that in the included studies, the biomarkers are always compared to the clinical diagnosis of epilepsy by a physician. We also added a planned comparison of the automated markers with the visual identification of IEDs on EEG in the **Introduction** and **Methods** section:

"We will perform a systematic review of diagnostic test accuracy for automated methods of interictal EEG analysis to distinguish between patients with and without epilepsy without relying on the detection of spikes. The questions that this review addresses are the following: What is the current evidence on the performances of automatically extracted hidden markers compared to the clinical diagnosis of epilepsy by a physician? […]" (pp. 5–6, ll. 84–92)
"The interictal EEG is key in informing the diagnosis of epilepsy, solely based on the visual identification of interictal spikes.[42] Despite years of research on computational biomarkers of epilepsy, only these spikes are currently used in clinical settings.[1,17,18] This review aims to systematically evaluate the performances of hidden interictal markers of epilepsy on EEG against the clinical diagnosis by a physician, describe the data processing pipelines favored by the researchers to classify the EEG for epilepsy diagnosis, and identify the pitfalls that prevent clinical translation of these algorithms." (pp. 12–13, ll. 256–263)
"The questions that this review addresses are the following: […] What is the benefit over the visual identification of IEDs on routine EEG? And what are the different algorithms that have been tested and how does their diagnostic accuracy compare?" (pp. 5–6, ll. 86–93)
"We will compare the performance of the computational markers for the diagnosis of epilepsy to the visual identification of IEDs on EEG." (p. 11, ll. 229–230)

Reviewer 3

The publication needs a major revision.

Comment 1: *Clearly define the research gap and highlight the outcomes.*

Authors' response: The research gap is described in the **Introduction**. As per the suggestion of the reviewer, we modified this passage to better emphasis the requirement for a systematic review addressing this research question:

"Decades of research have suggested that the automated analysis of EEG can identify hidden differences between with epilepsy and non-epileptic subjects in terms of connectivity[22–24], signal predictability and complexity[25,26], spectral power[27,28], and chaoticity[29]. Computational analysis of EEG holds the promise of extracting information that is invisible to the naked eye of the human interpreter, in an objective and reproducible manner. Discovering new, non-visible markers of epilepsy could increase the diagnostic yield of the EEG, improve its accessibility, and reduce costs, especially in settings where the expertise of a fellowship-trained neurophysiologist is unavailable[18,30]. In spite of this, none of the proposed non-visible markers of epilepsy have made it into clinical practice[10,31]. This discrepancy calls attention to the lack of comprehensible and systematic evaluation of these new methods." (p. 5, ll. 74–83)

The outcome of the study is the diagnostic accuracy of computational markers on routine EEG for the diagnosis of epilepsy: we believe that this is highlighted in the **Title**, **Introduction** (ll. 84–93), **Methods: Analyses** (ll. 219–230), and **Discussion** (ll. 256–263).

Comment 2: *All the findings need to address and explained in well-formatted figures, tables, and graphs.*

Authors' response: The findings of the systematic review will be reported using figures and tables. As suggested by the reviewer, we added two planned figures: one to represent the sensitivity and

specificity in the ROC plane and another to visualize the evaluation of the risk of bias. We also added a planned table for the description of the EEG processing pipelines used in the studies. We describe the exact figures and tables that will be provided in the systematic review in the **Methods: Data synthesis/Analyses sections**:

> "We will provide a table summarizing every published study included in the review, comparing the studies' design, population, reference standard, dataset size, data processing methods, and diagnostic accuracy. We will also provide a figure that summarizes the risk of bias for each item in the adapted QUADAS-2 tool, comparing 1) every individual article included in the review, and 2) every public dataset that is used in ≥ 2 studies." (p.10, ll. 204–206)
> "We will summarize in a table the methods used by the different articles during the pipeline's algorithm (pre-processing, feature extraction, feature selection, and classification algorithm), along with the proportion of studies that used each method." (p. 11, ll. 215–217)
> "We will present the results of the analyses with forest plots." (p. 11, ll. 228–229)
> "Reporting bias for sensitivity and specificity will be evaluated by visual inspection of funnel plots." (p.12, l. 252)
> "For heterogeneity in the ROC plane, we will compute the area of the 95% prediction ellipse and present the results on a scatterplot in the ROC plane.[39]" (p. 12, ll. 235–236)

As this is a protocol for a systematic review, there are no results to present. However, we have included the PRISMA checklist as Supplementary material.

VERSION 2 – REVIEW

REVIEWER	Cynthia Sharpe Starship Children's Health, Neurology
REVIEW RETURNED	20-Dec-2022

GENERAL COMMENTS	This will be an extremely useful review