## 1.6 Overall synopsis of the clinical investigation

| | |
|---|---|
| Title | Performance of BIOSERENITY AI EEG normal/abnormal algorithm in classifying normal - abnormal EEG automatically on real-world data (RWD). |
| Short title | BIOSERENITY AI normal/abnormal EEG post-market study. |
| Lay title, if applicable | Not applicable |
| CIP version and date | Version 1.0– 06/10/2026 |
| EUDAMED Single Reference Number (SRN) or CIV-ID, if previously assigned | FR-MF-000039391 |
| CI modification number, if applicable | Version 1.0 |
| Sponsor name and address | BIOSERENITY MEDICAL DEVICES GROUP<br>6/8 rue Jean Antoine de Baïf,<br>75013 Paris |
| Participating Location(s) and country(ies) | Data will be collected from several hospitals in France through Neurophy, BioSerenity tele-interpretation service. |
| Name of Investigational Device | BIOSERENITY AI EEG NEURO -AI EEG NORMAL/ABNORMAL |
| Clinical investigation Purpose and Background<br>• Rationale for CI<br>• Background of device and condition<br>• Current standard of care | An electroencephalogram (EEG) exam, in conjunction with clinical evaluation, is frequently used in neurological investigations but requires very specific expertise.<br><br>Because of this need for specific expertise, Artificial Intelligence (AI) can help with diagnosis and democratize the use of EEGs. Several papers have been published to address this classification challenge using Artificial Intelligence. Deep-learning approaches have repeatedly demonstrated expert-level accuracy in discriminating normal from abnormal recordings (van Leeuwen et al., 2019; Gemein et al., 2020), and a comprehensive AI model has recently been reported to interpret routine clinical EEGs at a level comparable to human experts (Tveit et al., 2023). However, the performance of such models is most often established on retrospective, curated or single-source datasets, and frequently decreases when the models are applied to external populations, centers and equipment; to date only a minority of AI tools have successfully transitioned into routine clinical practice (Mansilla et al., 2024).<br><br>The BIOSERENITY AI normal/abnormal EEG is an advanced algorithm designed to automate the classification of EEG records as normal or abnormal.<br><br>Performance and safety of BIOSERENITY AI normal/abnormal EEG were demonstrated in a clinical validation retrospective study (Health Data Hub n° 23976707), ensuring they meet or exceed the standards set by currently marketed devices. On the basis of these results, the device obtained its CE mark |

| | |
|---|---|
| | under Regulation (EU) 2017/745 and has been placed on the market since December 2025.<br><br>In accordance with the post-market surveillance (PMS) and post-market clinical follow-up (PMCF) obligations defined in Articles 83 to 86 and Annex XIV, Part B of Regulation (EU) 2017/745, the safety and performance of the device must now be confirmed under its actual conditions of use.<br><br>The present study has therefore been designed to evaluate the algorithm on real-world data (RWD), i.e. EEG recordings acquired in routine clinical practice across several French hospitals through the Neurophy tele-interpretation service.<br><br>The initial validation study established performance on a controlled, retrospective dataset, whereas real-world recordings differ in case mix, indication spectrum, acquisition conditions, signal quality and operator practices. Confirming that the algorithm maintains its claimed level of performance on this more heterogeneous and representative population is essential both to detect any performance drift after market introduction and to substantiate the generalizability of the device, which is a recognized critical issue for medical AI (Mansilla et al., 2024).<br><br>Accordingly, in order to confirm the consistency and robustness of the algorithm in routine this post-market study will benchmark the real-world performance of the BIOSERENITY AI normal/abnormal EEG algorithm against the performance claimed at the time of CE marking, using the score of a qualified physician as the reference standard, and consecutively the performance reported for a comparable marketed device. |
| Study device and Comparator | The clinical validation of the algorithms will be based on its performance to classify the EEG records in 2 categories:<br>• **"IA 1" for a "certainly normal" trace:**<br>o absence of pathological activity<br>o normal background activity with alpha rhythm<br>o possibility of drug impregnation (e.g. benzodiazepines), but the reading leads to the conclusion that the trace is normal. The presence of a rapid rhythm does not prevent conclusion.<br>o Possibility of sedation trace, but the reading allows to conclude that the trace is normal.<br>• **"IA 3" for a "certainly abnormal" trace.** If the trace falls into one of the following categories, the physicians can specify it:<br>o IA3L: Lesional (focal pathological activity, indication for imagery).<br>o IA3E: Diffuse encephalopathy.<br>o IA3S: Status epilepticus (focal or generalized).<br>o IA3A: Other/undetermined.<br>o IA3Ep: Presence of epileptiform activity<br>Else, the trace is only annotated IA3. |

| | |
|---|---|
| | This algorithm classification will be compared with the scoring provided by a qualified physician who has scored the records in routine clinical practice. The physician score will be used as a reference to assess algorithm performance.<br><br>For physicians, a third category of EEG recordings grading **"IA 2"** exists when it is impossible to decide between certainly normal and certainly abnormal (exceptional notation). It is added in case of:<br><br>• presence of too many artifacts<br>• drug impregnation or sedation preventing any conclusion (excessive presence of rapid rhythm) |
| Clinical development stage | Post-market study |
| Design of the clinical investigation | Retrospective, non-interventional, multicenter, post-market study |
| Objectives | <u>Primary objective</u><br>• To evaluate the classification performance of BIOSERENITY AI normal/abnormal EEG algorithm on real-world data against the score of physicians versus the performance claimed at the time of CE marking (initial validation study).<br><u>Secondary objectives</u><br>• To evaluate the classification performance of BIOSERENITY AI normal/abnormal EEG algorithm on real-world data against the score of a similar device (autoSCORE).<br>• To evaluate the classification performance of the algorithm according to patients' demographic data. |
| Primary endpoints | Evaluation of sensibility, specificity, balanced accuracy, positive and negative predictive values of the classification of all the records on the scores of physicians versus the performance claimed at the time of CE marking using the score of a qualified physician as the reference standard.<br>• Sensitivity (true positive rate) is the probability of a positive test result, conditioned on the individual truly being positive.<br>• Specificity (true negative rate) is the probability of a negative test result, conditioned on the individual truly being negative.<br>• Balanced accuracy is the average of the recall obtained in each class.<br>• Positive predictive value is a measure of how often a positive test result is actually positive<br>• Negative predictive value is a measure of how often a negative test result is actually negative |
| Secondary endpoints | • Evaluation of classification performance of the algorithm on real-world data against the score of a similar device (autoSCORE) on Sensitivity, Specificity, Balanced accuracy, Positive predictive value and Negative predictive value .<br>• Evaluation of classification performance of the algorithm on patients' demographic data. |
| Description of participants / study population | The study population will include adult patients who performed an EEG for neurological clinical examination for several indications as epilepsy, dementia, coma, stroke. |

| | |
|---|---|
| | Around 2500 patients will be included in the study in order to obtain 2500 EEG records exploitable. |
| Inclusion criteria | 1. ≥18 years<br>2. Male or female<br>3. Having completed a standard EEG record (approx. 20min recording)<br>4. Informed and non-opposed for the use of their personal data for the conduct of the research<br>5. Neuronaute tracings only (not necessarily Icecap only)<br>6. Tracings for which the interpreter has indicated, after visual analysis of the EEG trace, one of the codes “IA 1”, “IA2” or “IA 3”. |
| Exclusion criteria | 1. Tracings for which the interpreter has not indicated the code “IA 1”, “IA2” or “IA 3”.<br>2. EEG lacking the 19 channels of the 10/20 positioning or for which some channels are non-functional / too artifactual. |
| Sample size | The objective of the study is to evaluate the performance of the study device in detecting and classifying EEG as normal or abnormal on real-world data and to compare it against the performance obtained in the initial validation study.<br><br>Due to the great heterogeneity of pathologies and populations that will be studied and analyzed, around 2500 EEG records will be required. This sample size is aligned with the initial validation study to allow a like-for-like comparison. |
| Duration and follow up of the clinical investigation | The target period for data collection will span from 16 June 2026 to 02 October 2026. |
| Statistical considerations | **Primary endpoint analysis**<br>Computation of performance metrics will be performed by considering the results of a physician as reference, and 95% confidence intervals (CI) are computed by bootstrap resampling (1000 resamples).<br><br>Results will be compared to the performance claimed at the time of CE marking using the score of a qualified physician as the reference standard with a 10% marge.<br><br>**Secondary endpoint analyses**<br>• Computation of performance metrics will be performed by considering the results compared to the performance reported for a comparable marketed device, autoSCORE / SCORE-AI.<br>• Computation of performance metrics will be performed by considering the results for each demographic data parameter. |