



January 7, 2024

Holberg EEG AS
Smriti Franklin
QA/RA Manager
Fjøsangerveien 70A
5068 Bergen, Norway

Re: K231068
Trade/Device Name: autoSCORE
Regulation Number: 21 CFR 882.1400
Regulation Name: Electroencephalograph
Regulatory Class: Class II
Product Code: OMB
Dated: December 8, 2023
Received: December 8, 2023

Dear Smriti Franklin:

We have reviewed your section 510(k) premarket notification of intent to market the device referenced above and have determined the device is substantially equivalent (for the indications for use stated in the enclosure) to legally marketed predicate devices marketed in interstate commerce prior to May 28, 1976, the enactment date of the Medical Device Amendments, or to devices that have been reclassified in accordance with the provisions of the Federal Food, Drug, and Cosmetic Act (the Act) that do not require approval of a premarket approval application (PMA). You may, therefore, market the device, subject to the general controls provisions of the Act. Although this letter refers to your product as a device, please be aware that some cleared products may instead be combination products. The 510(k) Premarket Notification Database available at <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm> identifies combination product submissions. The general controls provisions of the Act include requirements for annual registration, listing of devices, good manufacturing practice, labeling, and prohibitions against misbranding and adulteration. Please note: CDRH does not evaluate information related to contract liability warranties. We remind you, however, that device labeling must be truthful and not misleading.

If your device is classified (see above) into either class II (Special Controls) or class III (PMA), it may be subject to additional controls. Existing major regulations affecting your device can be found in the Code of Federal Regulations, Title 21, Parts 800 to 898. In addition, FDA may publish further announcements concerning your device in the Federal Register.

Additional information about changes that may require a new premarket notification are provided in the FDA guidance documents entitled "Deciding When to Submit a 510(k) for a Change to an Existing Device" (<https://www.fda.gov/media/99812/download>) and "Deciding When to Submit a 510(k) for a Software Change to an Existing Device" (<https://www.fda.gov/media/99785/download>).

Your device is also subject to, among other requirements, the Quality System (QS) regulation (21 CFR Part 820), which includes, but is not limited to, 21 CFR 820.30, Design controls; 21 CFR 820.90, Nonconforming product; and 21 CFR 820.100, Corrective and preventive action. Please note that regardless of whether a change requires premarket review, the QS regulation requires device manufacturers to review and approve changes to device design and production (21 CFR 820.30 and 21 CFR 820.70) and document changes and approvals in the device master record (21 CFR 820.181).

Please be advised that FDA's issuance of a substantial equivalence determination does not mean that FDA has made a determination that your device complies with other requirements of the Act or any Federal statutes and regulations administered by other Federal agencies. You must comply with all the Act's requirements, including, but not limited to: registration and listing (21 CFR Part 807); labeling (21 CFR Part 801); medical device reporting (reporting of medical device-related adverse events) (21 CFR Part 803) for devices or postmarketing safety reporting (21 CFR Part 4, Subpart B) for combination products (see <https://www.fda.gov/combination-products/guidance-regulatory-information/postmarketing-safety-reporting-combination-products>); good manufacturing practice requirements as set forth in the quality systems (QS) regulation (21 CFR Part 820) for devices or current good manufacturing practices (21 CFR Part 4, Subpart A) for combination products; and, if applicable, the electronic product radiation control provisions (Sections 531-542 of the Act); 21 CFR Parts 1000-1050.

Also, please note the regulation entitled, "Misbranding by reference to premarket notification" (21 CFR 807.97). For questions regarding the reporting of adverse events under the MDR regulation (21 CFR Part 803), please go to <https://www.fda.gov/medical-devices/medical-device-safety/medical-device-reporting-mdr-how-report-medical-device-problems>.

For comprehensive regulatory information about medical devices and radiation-emitting products, including information about labeling regulations, please see Device Advice (<https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance>) and CDRH Learn (<https://www.fda.gov/training-and-continuing-education/cdrh-learn>). Additionally, you may contact the Division of Industry and Consumer Education (DICE) to ask a question about a specific regulatory topic. See the DICE website (<https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/contact-us-division-industry-and-consumer-education-dice>) for more information or contact DICE by email (DICE@fda.hhs.gov) or phone (1-800-638-2041 or 301-796-7100).

Sincerely,

Jay R. Gupta -S

Jay Gupta
Assistant Director
DHT5A: Division of Neurosurgical,
Neurointerventional
and Neurodiagnostic Devices
OHT5: Office of Neurological
and Physical Medicine Devices

Office of Product Evaluation and Quality
Center for Devices and Radiological Health

Enclosure

Indications for Use

510(k) Number (if known)
K231068

Device Name
autoSCORE

Indications for Use (Describe)

1. autoSCORE is intended for the review, monitoring and analysis of EEG recordings made by electroencephalogram (EEG) devices using scalp electrodes and to aid neurologists in the assessment of EEG. This device is intended to be used by qualified medical practitioners who will exercise professional judgment in using the information.
2. The spike detection component of autoSCORE is intended to mark previously acquired sections of the patient's EEG recordings that may correspond to spikes, in order to assist qualified clinical practitioners in the assessment of EEG traces. The spike detection component is intended to be used in patients at least three months old. The autoSCORE component has not been assessed for intracranial recordings.
3. autoSCORE is intended to assess the probability that previously acquired sections of EEG recordings contain abnormalities, and classifies these into pre-defined types of abnormalities, including epileptiform and non-epileptiform abnormalities. autoSCORE does not have a user interface. autoSCORE sends this information to the EEG reviewing software to indicate where markers indicating abnormality are to be placed in the EEG. autoSCORE also provides the probability that EEG recordings include abnormalities and the type of abnormalities. The user is required to review the EEG and exercise their clinical judgement to independently make a conclusion supporting or not supporting brain disease.
4. This device does not provide any diagnostic conclusion about the patient's condition to the user. The device is not intended to detect or classify seizures.

Type of Use (Select one or both, as applicable)

Prescription Use (Part 21 CFR 801 Subpart D)

Over-The-Counter Use (21 CFR 801 Subpart C)

CONTINUE ON A SEPARATE PAGE IF NEEDED.

This section applies only to requirements of the Paperwork Reduction Act of 1995.

DO NOT SEND YOUR COMPLETED FORM TO THE PRA STAFF EMAIL ADDRESS BELOW.

The burden time for this collection of information is estimated to average 79 hours per response, including the time to review instructions, search existing data sources, gather and maintain the data needed and complete and review the collection of information. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden, to:

Department of Health and Human Services
Food and Drug Administration
Office of Chief Information Officer
Paperwork Reduction Act (PRA) Staff
PRASStaff@fda.hhs.gov

"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB number."

510(K) Summary

1. SUBMITTER

Holberg EEG AS

Fjøsangerveien 70A

5068 Bergen, Norway

Phone: +47 926 44 261

Contact Person: Smriti Franklin

Date Prepared: March 23rd, 2023

2. DEVICE IDENTIFICATION

Trade Name: autoSCORE

Common Name: Automatic event detection software for full-montage electroencephalograph

Classification Name and Regulation Number: Electroencephalograph, 21 CFR 882.1400

Regulatory Class: II

Product Code: OMB

3. PREDICATE DEVICES

Primary Predicate Device

Trade/Device Name: encevis, K171720

Additional Predicate Device

Trade/Device Name: Persyst 13, K151929

4. DEVICE DESCRIPTION

autoSCORE is a software-only decision support product intended to be used with compatible electroencephalography (EEG) review software. It is intended to assist the user when reviewing EEG recordings, by assessing the probability that previously acquired sections of EEG recordings contain abnormalities, and classifying these into pre-defined types of abnormality. autoSCORE sends this information to the EEG software to indicate where markers indicating abnormality are to be placed in the EEG. autoSCORE uses an algorithm that has been trained with standard deep learning principles using a large training dataset.

autoSCORE also provides an overview of the probability that EEG recordings and sections of EEG recordings include abnormalities, and which type(s) of abnormality they include. This is performed by identifying spikes of epileptiform abnormalities (Focal epileptiform and Generalized epileptiform) as well identifying non-epileptiform abnormalities (Focal Non-epileptiform and Diffuse Non-epileptiform).

The user is required to review the EEG and exercise their clinical judgement to independently make a conclusion supporting or not supporting brain disease.

autoSCORE cannot detect or classify seizures. The recorded EEG activity is not altered by the information provided by autoSCORE. autoSCORE is not intended to provide information for diagnosis but to assist clinical workflow when using the EEG software.

5. INDICATIONS FOR USE/ INTENDED USE

5.1 INDICATIONS FOR USE STATEMENT

1. autoSCORE is intended for the review, monitoring and analysis of EEG recordings made by electroencephalogram (EEG) devices using scalp electrodes and to aid neurologists in the assessment of EEG. This device is intended to be used by qualified medical practitioners who will exercise professional judgment in using the information.
2. The spike detection component of autoSCORE is intended to mark previously acquired sections of the patient's EEG recordings that may correspond to spikes, in order to assist qualified clinical practitioners in the assessment of EEG traces. The spike detection component is intended to be used in patients at least three months old. The autoSCORE component has not been assessed for intracranial recordings.
3. autoSCORE is intended to assess the probability that previously acquired sections of EEG recordings contain abnormalities and classifies these into pre-defined types of

abnormalities, including epileptiform and non-epileptiform abnormalities. autoSCORE does not have a user interface. autoSCORE sends this information to the EEG reviewing software to indicate where markers indicating abnormality are to be placed in the EEG. autoSCORE also provides the probability that EEG recordings include abnormalities, and the type of abnormalities. The user is required to review the EEG and exercise their clinical judgement to independently make a conclusion supporting or not supporting brain disease.

4. This device does not provide any diagnostic conclusion about the patient's condition to the user. The device is not intended to detect or classify seizures.

5.2 INTENDED USE ENVIRONMENT

autoSCORE is intended to be used in environments where clinical EEGs are acquired or reviewed by suitably trained and qualified professionals.

autoSCORE is intended to be used for the analysis of EEG that has been recorded in environments suitable for adult and pediatric routine EEG acquisition according to best clinical practice, excluding acquisition environments for ICU and neonatal recordings. autoSCORE is not validated for EEG recorded in a home/ambulatory environment or any non-hospital/EEG laboratory setting.

5.3 INTENDED PATIENT POPULATION

autoSCORE use is restricted to EEG recordings from patients over 3 months of age. autoSCORE cannot be used for EEG recordings from neonatal patients. This restriction applies to all features of autoSCORE. There are no other restrictions regarding the patient population.

6. SUBSTANTIAL EQUIVALENCE DISCUSSION

The following table 1 compares autoSCORE to the predicate device with respect to intended use, technological characteristics and operating principles. The comments section provides further information on the determination of substantial equivalence.

Table 1: Comparison of autoSCORE against predicate devices.

		autoSCORE	encevis	Persyst 13	Comments
Device, regulation and sponsor details	510k Reference	Subject device	K171720	K151929	N/A
	Product Code	OMB	OMB	OMB	Identical
	Class	II	II	II	Identical
	Regulation Number	21 CFR 882.1400	21 CFR 882.1400	21 CFR 882.1400	Identical
	Regulation Name	Electroencephalograph	Electroencephalograph	Electroencephalograph	Identical

		autoSCORE	encevis	Persyst 13	Comments
	Manufacturer	Holberg EEG AS	AIT Austrian Institute of Technology GmbH	Persyst Development Corporation	N/A
	Device Type	Software-only Device	Software-only Device	Software-only Device	Identical
Device Description and Features	General Device Description	EEG Review and Analysis Software	EEG Review and Analysis Software	EEG Review and Analysis Software	Identical
	Identifies Spikes	Yes	Yes	Yes	Identical
	Assessment and categorization of abnormalities including probability in previously acquired sections of EEG	Yes	No	No	Different
	Type of EEG	Scalp EEG	Scalp EEG	Scalp EEG	Identical
	Intended Use Environments	<p>autoSCORE is intended to be used in environments where clinical EEGs are acquired or reviewed by suitably trained and qualified professionals.</p> <p>autoSCORE is intended to be used for the analysis of EEG that has been recorded in environments suitable for adult and pediatric routine EEG acquisition according to best clinical practice, excluding acquisition environments for ICU and neonatal recordings. autoSCORE is not validated for EEG recorded in a home/ambulatory environment or any non-hospital/EEG laboratory setting.</p>	<p>encevis is intended to be used in environments where clinical EEGs are acquired or reviewed by suitably trained and qualified professionals.</p> <p>encevis Spike Detection component is intended to be used in adult patients greater than or equal to 18 years. encevis Spike Detection performance has not been assessed for intracranial recordings.</p>	<p>Persyst 13 is intended to be used in environments where clinical EEGs are acquired or reviewed by suitably trained and qualified professionals.</p> <p>The Spike Detection component is intended to be used in patients at least one month old. Persyst 13 Spike Detection performance has not been assessed for intracranial recordings.</p>	Similar. See Patient age for intended population comparison.
	Population age	> 3 months	Adults (age > 18 years)	> 1 month	Minimum patient age more than the predicate device.

	autoSCORE	encevis	Persyst 13	Comments	
Device Operation	Design Input	Raw EEG Signal	Raw EEG signal	-	Identical
	Design Input files	Calculation is based on EEG data recorded by external EEG systems. They are read from the EEG data provided by the EEG system	Calculation is based on EEG data recorded by external EEG systems. They are either read from the EEG file provided by the EEG system or can be sent to encevis using the interface provided by AIT (AITInterfaceDLL)	Calculation is based on EEG data recorded by external EEG systems. They are read from the EEG file provided by the EEG system	Identical (No AIT interface)
	Algorithm	Convolutional Neural Network	Convolutional Neural Network	Neural Network	Identical
	User-defined parameters	No parameters in spike detection algorithm can be changed by the user	No parameters in spike detection algorithm can be changed by the user	-	Similar
	Type of EEG-Analysis	Post-hoc analysis	Post-hoc analysis	Post-hoc analysis	Identical
	Design Output	Spike Detection component makes the results available to the user in form of markers	Spike Detection component makes the results available to the user in form of markers	Spike Detection component makes the results available to the user in form of markers	Identical
	Device Outputs	Identification and categorization of epileptiform and non-epileptiform abnormalities including probability that EEG recordings include abnormalities, and the type of abnormalities. These outputs are given at both recording Level and marker Level.	Identification of epileptiform abnormalities (spikes). These outputs are given at marker level.	Identification of epileptiform abnormalities (spikes). These outputs are given at marker level.	Some predicate device features are not included in autoSCORE. These include seizure detection, burst suppression, aEEG, rhythmic and periodic patterns and frequency bands.
	Output Files	Results are returned back to the host software after analysis.	Results are stored in a database and/or sent over the interface AITInterfaceDLL to an external EEG system. User output is given by graphical user interfaces.	Results are stored in additional files in the file system placed in the same folder as the EEG file. User output is given by graphical user interfaces.	Similar

Diagnostic conclusion	This device does not provide any diagnostic conclusion about the patient's condition to the user.	This device does not provide any diagnostic conclusion about the patient's condition to the user.	This device does not provide any diagnostic conclusion about the patient's condition to the user.	Identical
	autoSCORE	encevis	Persyst 13	Comments
User	This device is intended to be used by qualified medical practitioners who will exercise professional judgment in using the information.	This device is intended to be used by qualified medical practitioners who will exercise professional judgment in using the information.	This device is intended to be used by qualified medical practitioners who will exercise professional judgment in using the information.	Identical
Compatible and interoperable Equipment and software	autoSCORE can read and process EEG data from Natus® NeuroWorks® software	encevis can read and process EEG data from several EEG vendors. A list of compatible EEG systems can be found on http://www.encevis.com	Persyst 13 can read and process EEG data from several EEG vendors. A list of compatible EEG systems can be found on http://www.persyst.com/support/supported-formats/	Similar

Color Key			
Identical/Similar Characteristics		Different or N/A Characteristics	

Comparison of Intended Use/ Indications for Use

The Indications for Use statement for autoSCORE is similar to the predicate devices. However, autoSCORE does not contain certain predicate device features including seizure detection, burst suppression, and other quantitative measures. Indications for use statement point 1, 2, and 4 are identical to the respective parts of predicate devices indications for use statement. Point 3 of the indications for use statement describes autoSCORE's technological characteristics, including additional outputs that are different from the predicate devices.

These differences do not alter the intended use of the device, nor do they affect the safety and effectiveness of the device relative to the predicates. Both the subject and predicate devices have the same intended use for analyzing electroencephalograph data, identifying events including spike detection and producing outputs based on analysis of EEG for interpretation by a qualified user.

Comparison of Technological Characteristics

Technological differences between the subject and predicate devices have been highlighted in Table 1 above. There are additional features in the predicate devices, including seizure detection, analysis of additional quantitative features, and a different user interface, which

are outside the intended use of the subject device. These features are completely independent functions and do not affect the abnormality detection features of the subject device.

Both the predicate devices and subject device detect features related to epileptiform abnormalities (e.g. spikes). In addition to detecting epileptiform abnormalities, the subject device also detects non-epileptiform abnormalities. The subject device also provides the probability of the detected abnormality being an epileptiform abnormality, such as a focal epileptiform or generalized epileptiform abnormality, or a non-epileptiform abnormality, such as a focal non-epileptiform or diffuse non-epileptiform abnormality. The identification of additional abnormalities and categorization of these abnormalities does not affect the intended use of the device and does not pose any additional risks as compared to the predicate devices as evidenced through performance validation.

7. Performance Validation

Performance validation to evaluate autoSCORE performance was conducted in two parts:

- Non-Clinical Validation – To validate autoSCORE outputs against defined autoSCORE design inputs and user requirements.
- Clinical Validation – To validate autoSCORE performance against independent human experts and predicate devices.

These validations have been summarized below.

7.1 Non-clinical Performance Validation

Software verification and validation testing was conducted and documented in accordance with 2005 FDA Guidance, *Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices*.

Product Design and Software Requirements Traceability has been documented and verified against verification and validation test results.

Software verification and validation testing included:

1. Code Review
2. Unit level testing
3. System level testing
4. Integration level testing

The software for this device is determined as a “moderate” level of concern because a failure or latent flaw could indirectly result in minor injury to the patient or operator through incorrect information or through the action of a care provider.

Software verification and validation activities demonstrated that the device software meets

all software requirements.

7.2 Clinical Performance Validation

7.2.1 Clinical Performance Evaluation

A retrospective non-interventional comprehensive clinical validation was performed using de-identified data to evaluate performance of all autoSCORE features against Human Experts and predicate devices to establish substantial equivalence.

The following performance data have been provided in support of the substantial equivalence determination.

Table 2: Type of software performance test per feature. autoSCORE indicates the EEG as normal if it does not contain epileptiform or non-epileptiform abnormalities, and abnormal if it contains one or both of these abnormalities. Part 1, Part 2, and Part 5 of the clinical study show comparable results against Human experts where an EEG is marked as 'normal' or 'abnormal' by autoSCORE. Part 3 and 4 of the study include the assessment of presence and absence of epileptiform abnormalities in predicate devices and autoSCORE that feeds into the assessment of a normal or abnormal EEG.

Validation Tests Performed	autoSCORE Features - Identification and categorization of following abnormalities				
	Normal EEG	Spike Detection - epileptiform abnormalities		Non-epileptiform abnormalities	
		Focal epileptiform	Generalized epileptiform	Focal non-epileptiform	Diffuse non-epileptiform
Direct Comparison against predicate device	x	x	x	Not available in predicate	Not available in predicate
Benchmarking against both predicate devices with external gold standard EEGs	x	x	x	Not available in predicate	Not available in predicate
Comparison with Human Expert Evaluation	x	x	x	x	x

Clinical Performance Evaluations

For the performance evaluation of the autoSCORE spike detection device, the study was conducted to measure outputs of autoSCORE against the assessments from independent human experts as well as the spike detection from the predicate devices – encevis and Persyst 13.

Further, for the autoSCORE performance evaluation of additional technological features (not in the predicate device), the study was conducted to measure autoSCORE results of non-epileptiform abnormalities detection and categorization of abnormalities against Human Experts (HE).

The clinical validation study was carried out in five parts to compare the performance of autoSCORE with the human experts as well as with the predicate devices:

1. **Performance evaluation against human experts (single-Center):**
A single-center dataset of 4,850 EEGs assessed by 9 human experts assessing more than 1% of the EEGs each.
2. **Performance evaluation against human experts (multi-center):**
A multi-center dataset of 100 EEGs were assessed by 11 independent human experts.
3. **Direct comparison against primary predicate device (encevis):**
The same dataset of 100 EEGs used in Part 2 were used to evaluate performance against the primary predicate device, encevis.
4. **Benchmarking against primary and secondary predicate device (encevis and Persyst 13):**
A dataset of 58 EEGs was used to benchmark performance of both the primary predicate device encevis, the predicate device Persyst 13, and autoSCORE against human expert consensus.
5. **Performance evaluation against human experts (two centers):**
A hold-out dataset of 1315 EEGs not used for training of the AI model acquired from two centers were assessed by 15 human experts assessing more than 1% of the EEGs each.

The validation study was performed across five separate datasets with the following characteristics:

Table 3. Summary of study parts used for validation of autoSCORE.

Validation Parts	Number of sites	Sample size	Number of reviewers	Patient gender	EEG Duration min-max	Patient age min-max (median)	Pediatric (P) vs Adult (A)*

1. Performance evaluation against HE	1	4850	9	2527 (M) 2248 (F) 75 (unknown)	14 – 120 minutes	3 months – 106 years (39 years)	P – 1490 A – 3360
a. Recording level							
b. Marker Level							
2. Performance evaluation against HE	16	100	14	39 (M) 61 (F)	20 – 240 minutes	9 months – 95 years (26 years)	P – 43 A – 57
3. Direct comparison against predicate device							
4. Benchmarking against predicate devices and marker validation against HE consensus	4	58	3	27 (M) 31 (F)	20 – 30 minutes	2 – 77 years (36 years)	P – 13 A – 45
a. Recording level							
b. Marker Level							
5. Performance evaluation against HE	2	1315	15	636 (M) 642 (F) 37 (unknown)	14 - 240 minutes	3 months – 99 years (38 years)	P – 467 A – 848
a. Recording level							
b. Marker Level							

*Pediatric - <22 years; Adults -> 22 years

7.2.2 Study Population and Reference Standards

None of the EEGs used in the validation were used in the development of the AI model. The HEs providing the reference standards in the validation phase of Study 1, 2, 3, and 4 were different from those who participated in the development portion of the process.

Study 1 – The reference standard was based on 4850 EEGs described by multiple HEs, but a single HE reviewer per EEG. The HEs inserted markers in the EEGs defining if the EEG was abnormal or normal, and if abnormal, the abnormality categories, and served as reference standard both on recording level and marker level. The HE assessments were part of the routine EEG assessment in their respective hospitals, and the HEs had all relevant patient clinical information. Apart from age and gender, all clinical data was removed for this clinical validation to avoid any associated bias.

Study 2 – The reference standard was based on HE consensus of 11 HEs reviewing 100 EEGs. The HEs assessed if the EEGs on recording level were normal or abnormal, and if abnormal if the EEGs contained one or more of the abnormality categories Focal Epi, Gen Epi, Focal Non-Epi, and Diffuse Non-Epi. The HEs were blinded to all patient data except age and gender.

Study 3 – This study uses the same dataset as for study 2, and thus also the same HE consensus reference standard.

Study 4 – The reference standard was obtained by visual assessment of 58 EEGs by 3 HEs. Marker time points for the IEDs were recorded for each EEG and each HE. The reference standard was the majority consensus scoring of the HEs. This served as reference standard both on recording level and marker level for the IEDs. HEs were blinded to all patient data except age and gender.

Study 5 – The reference standard was based on 1315 EEGs described by multiple HEs, but a single HE reviewer per EEG. The HEs inserted markers in the EEGs defining if the EEG was abnormal or normal and, if abnormal, the abnormality categories, and served as reference standard both on recording level and marker level. The HE assessments were part of the routine EEG assessment in their respective hospitals, and the HEs had all relevant patient clinical information. Apart from age and gender, all clinical data was removed for this clinical validation to avoid any associated bias.

All relevant autoSCORE outputs were covered in the above studies.

7.2.3 Analytical Methods

The analytical methods used in this validation have been described in 7.2.3.1 and 7.2.3.2.

The figure 1 below shows a hierarchical representation of autoSCORE recording- and marker outputs and the conditions under which these outputs are presented in the Natus® NeuroWorks® user interface.

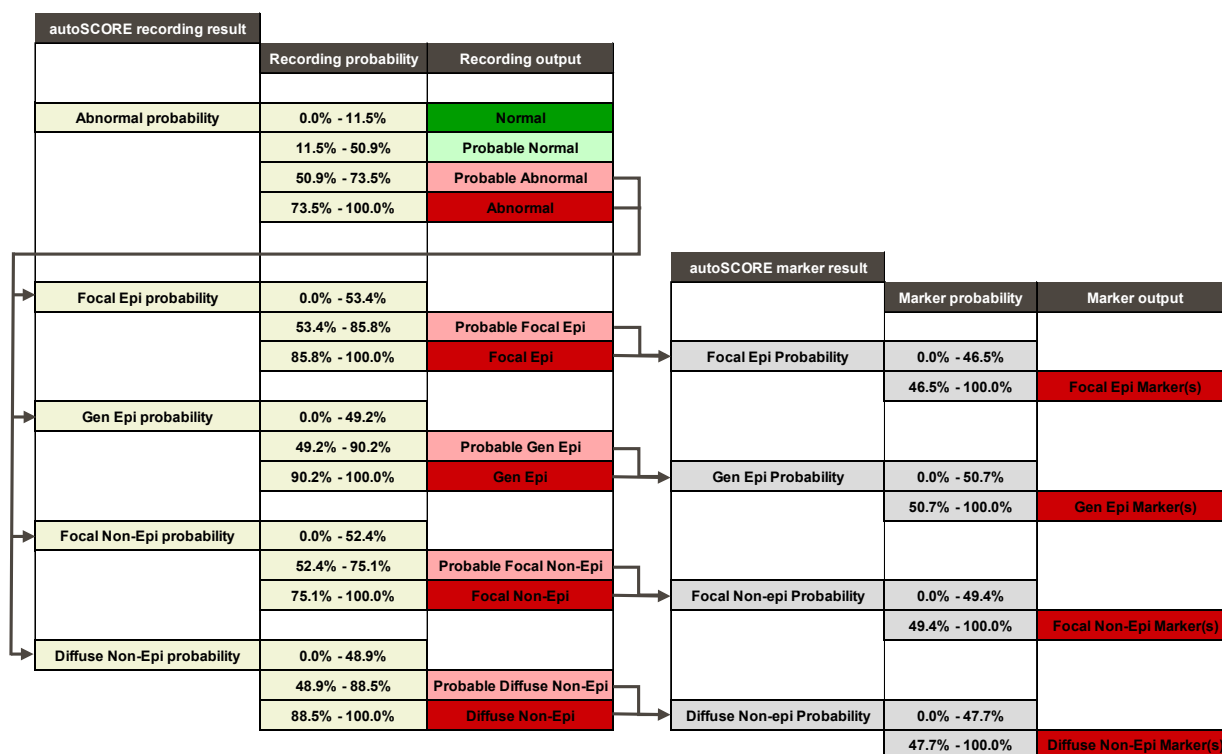


Figure 1: Hierarchical representation of autoSCORE recording and marker level outputs.

7.2.3.1 Comparison of performance with HEs

Binary Metrics

The binary metrics given in **Table 4**, **Table 5** and **Table 6** (sensitivity, specificity, PPV and NPV) in results section were computed independently for each study part and each feature (Normal/Abnormal, Focal Epi, Gen Epi, Focal Non-Epi, Diffuse Non-Epi) with 95% symmetric confidence intervals obtained using bootstrap resampling ($n \geq 10000$).

The following definitions were used for the binary metrics for the recording level outputs (where HE was used in study 1 and 5 and HE consensus in study 2):

TP – HE or HE consensus indicated that the condition is present and autoSCORE also indicates that the condition is present.

FP - HE or HE consensus indicated that the condition is not present but autoSCORE indicates that the condition is present.

TN - HE or HE consensus indicated that the condition is not present and autoSCORE also indicates that the condition is not present.

FN - HE or HE consensus indicated that the condition is present but autoSCORE indicates that the condition is not present.

For the marker level outputs the following definitions were used for study 1 and 5:

TP - A 16 second window where HE has indicated that an abnormality is present and autoSCORE also would place a marker for the same type of abnormality.

FP - A 16 second window where autoSCORE would place a marker for a specific type of abnormality and this window is either: 1) randomly extracted from an EEG that HE has

assessed as normal or 2) a 16 second window which the HE has assessed as not containing this type of abnormality.

FN- A 16 second window which HE has assessed as containing a specific type of abnormality where autoSCORE would not place a marker for this type of abnormality.

TN- A 16 second window where autoSCORE would not place a marker of a specific type of abnormality and this window is either: 1) randomly extracted from an EEG that the HE has assessed as normal or 2) a 16 second window which the HE has assessed as not containing this type of abnormality.

For the marker level outputs for study 4, validation was based on areas in the EEGs marked by autoSCORE and HE consensus markers. TP, TN, FP, and FN were derived from the resulting segmentation of the recording into areas marked only by autoSCORE, areas marked only by HE consensus and areas marked by both autoSCORE and HE consensus.

In the next step, values from the contingency matrices were used to calculate:

- Sensitivity, also referred to as True Positive Rate or $TPR = TP/(TP+FN)$
- Specificity, also referred to as True Negative Rate or $TNR = TN/(TN + FP)$
- $PPV = TP/(TP + FP)$
- $NPV = TN/(TN + FN)$
- Prevalence = $(TP + FN)/(TP + FP + TN + FN)$ = (number of true condition positive)/(number of samples).

Probability

To validate the probability output given by autoSCORE, several HE outputs were averaged in order to obtain a probability reference for the HEs. This grouping was done in different ways for study part 1/5 and for study part 2:

- Study part 1 and 5: The large number of EEG recordings allowed their grouping depending on autoSCORE probability values, applicable both for recording level and marker level. The grouping was uniform with 10 bins from 0% to 100%, each of 10 percent-points.
- Study part 2: The large number of HEs involved in the study allowed grouping EEGs depending on probability based on the number of HEs “voting” for the presence of the respective abnormalities, applicable only for recording level. Since each EEG was rated by 11 HEs, the granularity of this grouping was 9 percent-points.

The correlation coefficients given in **Table 4**, **Table 5** and **Table 6** and associated p values are Pearsons Correlation calculated using the python scipy stats package with the mean autoSCORE output and mean HE assessments in each bin as described above.

In study part 4, a similar discretization was performed on the markers placed. The ranges were here 100-90%, 90-80%, 80-70%, 70-60%, 60-50% and 50-0%. Only markers above threshold for user output were included, and a qualitative analysis of the number of overlapping markers with reference (HE consensus or primary predicate device) for each probability range.

Levels of abnormalities

Figure 2 below presents the strategy employed to validate categorical outputs. Categories in grey indicate areas that were not considered for determining TP, TN, FP and FN, and consequently, were not used for calculation performance parameters.

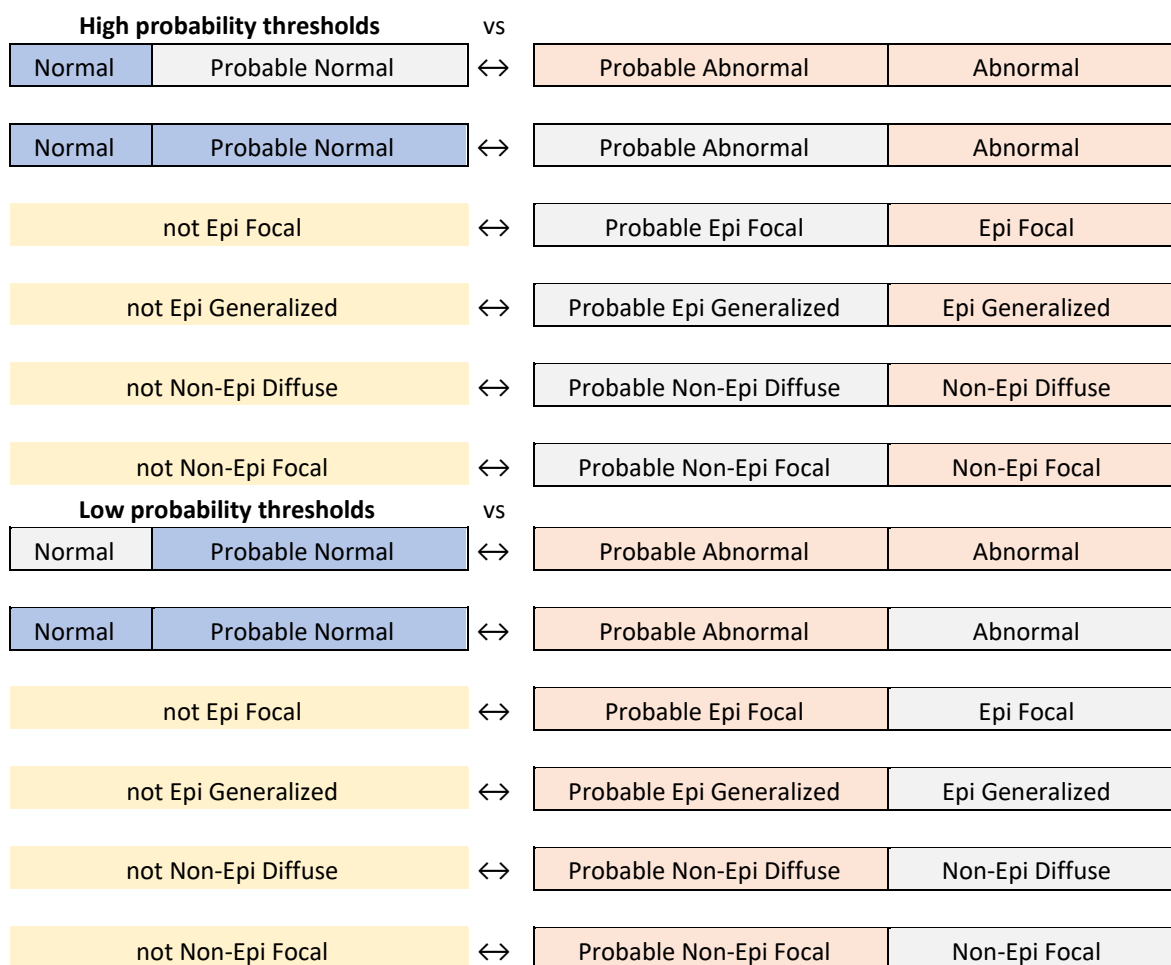


Figure 2: Schematic representation of strategy employed to validate the levels of abnormalities.

7.2.3.2 Comparison of performance with predicate devices

Binary Metrics – recording level

The binary metrics given in **Table 7** (accuracy, sensitivity, specificity, PPV and NPV) in the results section were computed in the same way as described above in the comparison with HE section. To allow comparison with predicate devices encevis (study part 3 and 4) and Persyst (study part 4), a number of assumptions and limitations had to be addressed:

- Only the epileptiform activity abnormalities can be validated against the predicate device. Focal Epi and Gen Epi parts of the autoSCORE output were merged.
- Since predicate devices are not designed to give recording level output, rules of interpretation had to be applied. If at least one spike is generated by the predicate device, the EEG recording is classified as Abnormal.

Probability level of abnormality – marker level

Only Focal Epi and Gen Epi markers placed by autoSCORE could be compared with the placement of the encevis spikes in study part 4.

7.2.4 Results of Performance Evaluation

7.2.4.1 Comparison of performance with HEs

Summary results obtained by the abovementioned methods are presented in Table 4 below. Tveit et al. [1] describes in detail the agreement between HEs and autoSCORE in comparison to HE-HE agreement.

Table 4: Performance results for autoSCORE classification of Abnormal EEG at recording-level based on comparison with HE assessment. Reference standards for each study part are discussed in a previous section of this document.

Age Group	Study Part	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Correlation coefficient (p-value)
All ages	Part 2 (n=100)	100 [100, 100]	88.4 [77.8, 97.4]	92.0 [84.5,98.3]	100 [100.0, 100.0]	0.96 (p < 10 ⁻⁵)
	Part 1 (n=4850)	83.1 [81.3, 84.8]	91.8 [90.8, 92.8]	84.9 [83.2, 86.6]	90.8 [89.8, 91.8]	0.99 (p < 10 ⁻⁵)
	Part 5 (n=1315)	87.8 [85.0, 90.5]	89.4 [87.2, 91.6]	86.0 [83.0, 88.8]	90.9 [88.8, 92.9]	0.99 (p < 10 ⁻⁵)
Adult	Part 2 (n=57)	100.0 [100.0, 100.0]	84.0 [68.0, 96.4]	88.9 [77.5, 97.4]	100.0 [100.0, 100.0]	0.95 (p < 10 ⁻⁵)
	Part 1 (n=3360)	85.0 [83.1, 86.8]	89.2 [87.8, 90.6]	86.2 [84.4, 88.0]	88.2 [86.7, 89.6]	0.98 (p < 10 ⁻⁵)
	Part 5 (n=848)	90.5 [87.6, 93.2]	84.6 [81.1, 87.9]	85.2 [81.9, 88.4]	90.1 [87.0, 92.9]	0.99 (p < 10 ⁻⁵)
Pediatric	Part 2 (n=43)	100.0 [100.0, 100.0]	94.4 [81.2, 100.0]	96.1 [87.1, 100.0]	100.0 [100.0, 100.0]	0.94 (p < 10 ⁻⁵)
	Part 1 (n=1490) ^A	71.6 [65.9, 77.1]	95.8 [94.6, 96.8]	76.3 [70.6, 81.8]	94.6 [93.4, 95.8]	0.96 (p < 10 ⁻⁵)
	Part 5 (n=467) ^B	79.7 [72.8, 86.2]	95.7 [93.4, 97.8]	88.7 [82.7, 94.0]	91.8 [88.9, 94.6]	0.95 (p < 10 ⁻⁵)

A results can be affected by low prevalence of 16%

B results can be affected by low prevalence of 29.6%

In addition to performance metrics results reported in Table 4, similar results were obtained to validate all categorical outputs (*Normal vs Probable Normal vs Probable Abnormal vs Abnormal*). The definition of the thresholds can be found in Figure 1 and the logic followed during validation of each category is presented in Figure 2 in 7.2.3 Performance evaluation measures. The categories shown in light grey in Figure 2 were omitted during validation of the respective performance metrics. It was shown that most of the autoSCORE results will be defined by the most confident Abnormal or Normal classification with PPV for all age groups in all studies equal to or higher than 90%. For the lower confidence levels Probable Abnormal and Probable Normal corresponding PPV are lower or equal, ranging from 80.3% to 100%.

Types of abnormality – Recording level

Summary results for the four types of abnormalities obtained by the abovementioned methods are presented in Table 5 below. Table 5 does not include results of the different age groups, but rather focuses on reporting the overall results (adult and pediatric patients combined) for each study part as no significant age-related differences were found. Tveit et al. [1] describes in detail the agreement between HE assessment and autoSCORE results in comparison to HE-HE agreement.

Table 5: Performance results for autoSCORE classification of abnormality types at recording-level based on comparison with HE assessment. Reference standards for each study part are discussed in a previous section of this document.

Abnormality Type	Study Part	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Prevalence	Correlation coefficient (p-value)
Focal Epi	Part 2 (n=100)	73.9 [54.5, 91.3]	88.3 [80.8, 94.9]	65.4 [45.8, 83.3]	91.9 [85.1, 97.4]	23.0%	0.85 (p = 0.00091)
	Part 1 (n=4850)	62.5 [55.9, 69.0]	95.0 [94.3, 95.6]	35.8 [30.9, 40.8]	98.3 [97.9, 98.6]	4.3 %	0.88 (p = 0.00089)
	Part 5 (n=1315)	66.5 [59.8, 73.1]	93.9 [92.5, 95.2]	64.8 [58.1, 71.4]	94.3 [92.9, 95.6]	14.5 %	0.96 (p < 10 ⁻⁵)
Gen Epi	Part 2 (n=100)	100.0 [100.0, 100.0]	94.1 [88.6, 98.8]	75.1 [54.5, 93.8]	100.0 [100.0, 100.0]	15,0 %	0.83 (p = 0.003)
	Part 1 (n=4850)	71.2 [62.3, 79.8]	98.3 [97.9, 98.6]	47.4 [39.7, 55.2]	99.4 [99.1, 99.6]	2.1 %	0.88 (p = 0.00067)
	Part 5 (n=848)	72.3 [62.0, 82.2]	96.1 [95.0, 97.2]	53.4 [43.8, 63.2]	98.3 [97.5, 98.9]	5.8 %	0.79 (p = 0.0066)
Diff Non-Epi	Part 2 (n=100)	87.5 [72.7, 100.0]	82.8 [74.0, 90.9]	61.7 [44.8, 78.0]	95.5 [89.7, 100.0]	24.0 %	0.93 (p < 10 ⁻⁵)
	Part 1 (n=4850)	65.2 [62.5, 67.9]	94.5 [93.7, 95.2]	79.1 [76.5, 81.6]	89.4 [88.4, 90.4]	24.3 %	0.98 (p < 10 ⁻⁵)
	Part 5 (n=467)	79.4 [74.5, 84.0]	94.0 [92.6, 95.4]	77.9 [73.0, 82.7]	94.5 [93.1, 95.8]	21.0 %	0.97 (p < 10 ⁻⁵)
Focal Non-Epi	Part 2 (n=100)	61.5 [42.1, 80.0]	93.2 [86.8, 98.6]	76.1 [56.2, 94.1]	87.4 [79.5, 94.1]	26.0 %	0.84 (p = 0.00056)
	Part 1 (n=4850)	65.2 [60.9, 69.5]	88.4 [87.4, 89.3]	38.3 [35.0, 41.7]	95.8 [95.2, 96.4]	10.0 %	0.97 (p < 10 ⁻⁵)
	Part 5 (n=467)	73.1 [67.4, 78.6]	89.6 [87.7, 91.4]	61.2 [55.6, 66.9]	93.7 [92.1, 95.1]	18.4 %	0.98 (p < 10 ⁻⁵)

In addition to performance metrics results reported in Table 5, similar results were obtained to validate all categorical outputs (*abnormality type vs Probable abnormality type*). The definition of thresholds can be found in Figure 1, and the logic followed during validation of each category is presented in Figure 2 in 7.2.3 Performance evaluation measures. The categories shown in light grey in Figure 2 were omitted during validation of the respective performance metrics. It was shown that most of the autoSCORE results will be defined by the most confident categories (Focal Epi, Generalized Epi, Diffuse Non-Epi and Focal Non-Epi) -with similar or greater PPV than for the less confident categories.

Types of abnormality – Marker level

Marker placement and correct type assignment were evaluated as part of Studies 1, 4, and 5. In study part 4, overlap of autoSCORE markers with areas indicated by HE

consensus as epileptiform discharges was validated. In study parts 1 and 5 it was validated if autoSCORE places markers in the same areas of EEG where previously HEs did. HEs also provided information about the type of abnormality assigned to the marker. Performance metrics calculated based on these results are provided in Table 6 together with the prevalence and correlation coefficient. autoSCORE places more markers in EEG recordings (declared as Abnormal on the recording level) than HEs, who typically select the most representative examples instead of all abnormality examples present in the recording. Consequently, the correlation coefficients and performance metrics values are lower for marker level abnormalities than recording level abnormalities leading to lower agreement between HEs and autoSCORE markers with assigned probability values below 90%. autoSCORE markers with high probability values indicate high agreement with markers inserted by HEs. Therefore, markers assigned with the highest level of confidence are most likely to indicate presence of each type of abnormality.

Table 6: Performance results for autoSCORE classification of abnormality types at the marker level based on comparison with HE assessment. Reference standards for each study part are discussed in a previous section of this document.							
Abnormality Type	Study Part	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Prevalence	Correlation coefficient (p-value)
Focal Epi + Generalized Epi	Part 4 (n=509)	87.3 [80.4, 92.7]	61.3 [56.1, 67.2]	38.9 [28.1, 50.4]	94.6 [92.1, 96.7]	22.0%	Qualitative not quantitative
Focal Epi	Part 1 (n=179700)	58.0 [57.0, 59.0]	98.0 [98.0, 98.1]	62.6 [61.6, 63.7]	97.6 [97.6, 97.7]	5.40%	0.88 (p = 0.0018)
	Part 5 (n=29989)	62.7 [61.1, 64.3]	96.6 [96.4, 96.9]	71.6 [70.1, 73.2]	95.1 [94.8, 95.3]	11.90%	0.95 (p < 10 ⁻⁵)
Generalized Epi	Part 1 (n=168441)	50.4 [49.2, 51.8]	99.6 [99.6, 99.6]	81.5 [80.2, 82.8]	98.3 [98.2, 98.4]	3.30%	0.81 (p = 0.0049)
	Part 5 (n=27521)	68.0 [65.3, 70.8]	99.4 [99.3, 99.5]	82.5 [79.9, 84.9]	98.7 [98.5, 98.8]	4.00%	0.92 (p = 0.00016)
Diffuse Non-Epi	Part 1 (n=204119)	52.7 [52.2, 53.2]	94.2 [94.1, 94.3]	69.6 [69.1, 70.2]	88.7 [88.5, 88.8]	20.20%	0.95 (p < 10 ⁻⁵)
	Part 5 (n=28986)	68.3 [66.5, 70.1]	94.3 [94.0, 94.6]	53.7 [52.0, 55.4]	96.8 [96.6, 97.1]	8.80%	0.89 (p = 0.0005)
Focal Non-Epi	Part 1 (n=179700)	63.0 [62.3, 63.8]	93.8 [93.7, 93.9]	51.3 [50.6, 52.0]	96.1 [96.0, 96.2]	9.40%	0.99 (p < 10 ⁻⁵)
	Part 5 (n=30168)	70.0 [68.5, 71.4]	92.5 [92.2, 92.8]	57.0 [55.6, 58.5]	95.6 [95.3, 95.8]	12.40%	0.93 (p = 0.00012)

7.2.4.2 Comparison of performance with predicate devices

Recording level

Summary results for study part 3 and 4 where autoSCORE's performance was compared with predicate devices are presented below in Table 7.

<p>Table 7: Performance results for autoSCORE classification of Focal Epi and Gen Epi EEG at the recording-level and predicate devices based on comparison with HE assessment. Reference standards</p>

for each study part are discussed in a previous section of this document.

Age Group	Study part	Device	Accuracy (%)	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
All ages	Part 3	autoSCORE	88.0 [81.0, 94.0]	90.0 [77.8, 100.0]	87.1 [78.8, 94.4]	95.3 [89.5, 100.0]	75.0 [60.0, 88.9]
		encevis	48.0 [38.0, 58.0]	96.7 [88.9, 100.0]	27.2 [17.1, 37.9]	95.0 [83.3, 100.0]	36.3 [25.9, 46.9]
	Part 4	autoSCORE	94.8 [87.9, 100.0]	93.3 [83.3, 100.0]	96.4 [88.0, 100.0]	93.1 [82.6, 100.0]	96.6 [88.5, 100.0]
		encevis	56.9 [44.8, 69.0]	100.0 [100.0, 100.0]	10.7 [0.0, 23.3]	100.0 [100.0, 100.0]	54.5 [41.1, 67.9]
		Persyst	53.5 [41.4, 65.5]	100.0 [100.0, 100.0]	3.6 [0.0, 12.0]	100.0 [100.0, 100.0]	52.6 [39.7, 65.5]
Adult	Part 3	autoSCORE	86.0 [77.2, 94.7]	84.6 [61.5, 100.0]	86.4 [75.6, 95.6]	95.0 [87.2, 100.0]	64.7 [40.0, 87.0]
		encevis	40.4 [28.1, 52.6]	92.3 [75.0, 100.0]	25.0 [12.8, 38.5]	91.6 [72.7, 100.0]	26.7 [14.3, 40.0]
	Part 4	autoSCORE	93.3 [84.4, 100.0]	92.6 [81.5, 100.0]	94.4 [81.2, 100.0]	89.5 [73.7, 100.0]	96.1 [87.5, 100.0]
		encevis	66.7 [53.3, 80.0]	100.0 [100.0, 100.0]	16.7 [0.0, 35.7]	100.0 [100.0, 100.0]	64.3 [50.0, 78.6]
Pediatric	Part 3	autoSCORE	90.7 [81.4, 97.7]	94.1 [80.0, 100.0]	88.5 [75.0, 100.0]	95.8 [86.2, 100.0]	84.2 [66.7, 100.0]
		encevis	58.2 [44.2, 72.1]	100.0 [100.0, 100.0]	30.8 [13.6, 50.0]	100.0 [100.0, 100.0]	48.6 [32.3, 65.0]
	Part 4	autoSCORE	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]
		encevis	23.1 [0.0, 46.2]	100.0 [100.0, 100.0]	0.0 [0.0, 0.0]		23.1 [0.0, 46.2]

Marker level

Overview of overlap of autoSCORE's Focal Epi and Gen Epi markers with encevis spikes dependent on probability assigned to the autoSCORE marker is presented below in Table 8. The highest agreement between both devices was obtained for markers with assigned higher probability. Therefore, markers assigned with the highest level of confidence are most likely to indicate presence of abnormality.

Age group	Probability range of autoSCORE markers	Number of autoSCORE markers	Number of autoSCORE markers overlapping with encevis spikes	Percentage of autoSCORE markers overlapping with encevis spikes
All	0 -50	15	2	13%

<i>ages</i>	50-60	57	11	19%
	60-70	59	19	32%
	70-80	112	37	33%
	80-90	93	50	54%
	90-100	173	149	86%
<i>Adult</i>	0 -50	11	1	9%
	50-60	46	8	17%
	60-70	47	17	36%
	70-80	79	29	37%
	80-90	72	38	53%
	90-100	150	128	85%
<i>All ages</i>	0 -50	4	1	25%
	50-60	11	3	27%
	60-70	12	2	17%
	70-80	33	8	24%
	80-90	21	12	57%
	90-100	23	21	91%

8. Biocompatibility, Electrical Safety, Electromagnetic Compatibility (EMC), and Mechanical Safety

autoSCORE is a software-only device. Biocompatibility, electrical safety, electromagnetic compatibility, and mechanical safety are not applicable to this device.

9. Statement of Substantial Equivalence

Since the predicate devices were cleared based in part on the results of clinical studies, and given the differences in device outputs, clinical testing was required to support substantial equivalence.

The non-clinical data support the safety of the device and the software verification and validation demonstrate that the autoSCORE device should perform as intended in the specified use conditions. The clinical data demonstrate that the subject device (autoSCORE) performs as well as the predicate devices that are currently marketed for the same intended use.

Therefore, autoSCORE is substantially equivalent to predicate devices in intended use. autoSCORE has some technological features that differ from predicate devices. Any differences between the subject and predicate device have no significant influence on safety or effectiveness. autoSCORE is at least as safe and effective as the legally marketed predicate devices, as established through performance testing. Therefore, the evidence provided demonstrates that autoSCORE is substantially equivalent to the predicate devices.

10. References

1. Tveit, J., et al., *Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence*. JAMA Neurology, 2023.