Responses to public comments
on the
Proposed clinical practice guideline for automated seizure detection using wearable devices

We would like to express our gratitude to those who submitted comments on the clinical practice guideline proposal on automated seizure detection using wearable devices. The comments helped us improve the manuscript. The draft was open for public comments for 50 days. We received 20 comments submitted via the homepage and three comments sent from representatives of the IBE. As there was considerable overlap between the comments, we have grouped them according to the topics and we answer them below.

The five phases of clinical validation studies

Concern was raised that the proposed definitions of the phases matched what was done in drug trials as opposed to those specific for algorithm studies. The definitions and standards we used were explicitly developed for automated seizure detection using wearable devices (Beniczky and Ryvlin, Epilepsia 2018). To facilitate understanding the level of evidence behind device validation studies, we proposed five phases that the broad audience is familiar with, like those using similar stratification in drug clinical trials. Phase 3 studies provide compelling evidence for the efficacy, based on studies with robust design (no or minimal potential for bias) and phase 4 are post-validation, in-field studies, focusing on the applicability in the home environment of the patients. The definition of each phase was based on the key features of evaluating the performance of wearable devices and seizure detection algorithms, including all the factors listed in the comment: label quality (reference standard), double blinding between labeler and developer of the algorithm, and independence of the test data from the training data (analysis and alarms) (Beniczky and Ryvlin, Epilepsia 2018).

More specifically, we will explain here, how we classified the three studies mentioned in the comments. The study by Halford et al. (Epilepsia, 2017) presenting the dataset used for the FDA clearance, did not have a pre-defined detection threshold and the analysis was not real time: the algorithm was run centrally on the dataset, after it has been recorded. Therefore, it does not meet the criteria of phase 3. Detection threshold is an essential part of the algorithms, as it directly affects the most important output parameters: sensitivity and false alarm rate. In the study by Halford et al., the threshold was fit post-hoc for the whole recorded dataset, to optimize the performance. However, that was not tested on an independent dataset, limiting its generalizability. Due to the lack of pre-defined threshold, the offline analysis, and the lack of independent

validation dataset, this study was classified as phase 2. The study by Onorati et al. (Epilepsia 2017) analyzed offline the recordings from three different wristbands to train and test two new machine learning classifiers and a published classifier. They used a cross-validation approach to construct and validate the classifiers. However, cross-validation techniques may provide too optimistic estimates for out-of-sample prediction performance, especially when study samples are relatively small. To qualify as phase 3 study (providing compelling evidence for the efficacy of the device) the pre-defined algorithm and detection threshold must run real-time (simultaneously with the recording) on the wearable device. Otherwise, the outcome is based on numerous assumptions, and there is a considerable risk of bias. Therefore, the study by Onorati et al. does not qualify as phase 3. A review paper by Regalia et al. (Epilepsy Res 2019) mentioned that a prospective study with a "fixed and frozen" algorithm was undertaken for the FDA clearance. However, that study has not been published in a peer-reviewed scientific journal yet. The study by Arends et al. (Neurology 2018) was classified as phase 4, because it was conducted in the home environment of the patients (in a residential care setting). The reference standard was based on video recordings of the seizures. Since the study addressed major convulsive seizures, video recordings provide sufficient information for the reference standard: experts can reliably identify generalized / bilateral tonic-clonic seizures based on video recordings. EEG is not necessary for reliable identification of these seizures. Because this study used real-time analysis of the signals recorded with the wearable device, with a pre-defined algorithm and pre-defined detection threshold, against a reference standard that was reliable for the addressed seizure type (video recordings of the major convulsive seizures), it also fulfils the criteria for phase 3 studies. It was mentioned in the comment that 14 patients from the training-data group appear in the test-data group. The paper specified that 14 participants were previously enrolled in a phase 1 study. However, in the phase 3-4 study, new datasets (new recordings, new seizures) were prospectively recorded in these patients too. Although it qualifies as phase 3, we agree that this was a limitation of that study, and we added that comment to the revised version.

Study inclusion

Several comments mentioned studies that we did not include. We would like to reiterate that we only included phase 2-4 studies, due to the high risk of bias in phase 0-1 studies. In the Discussion section of the revised manuscript we cite the pilot studies for the modalities that later on were validated in phase 3-4 studies (for example the study by Lockman et al., Epilepsy & Behavior 2011). However, including several hundreds of phase 0-1 studies on devices and algorithms that were

promising, but not properly validated, was beyond the scope and limitation of this ILAE-IFCN working group.

Automated seizure detection using sub-scalp devices for ultra-long-term EEG recordings have not been validated yet in phase 2-4 studies. In the study by Weisdorf et al. (Epilepsia, 2019), mentioned in one of the comments, the subcutaneous EEG recordings were reviewed visually to identify electrographic seizures. As our guideline specifically addresses automated detection, this study was clearly outside the scope of our work.

We are aware of two phase 2 studies published after the date last searched (October 16th, 2019), and therefore not included: van Westrhenen et al. (Epilepsia. 2020) and Vandecasteele et al. (Epilepsia 2020). The rigorous systematic review methodology required for this report imposed to stick to a definite final search date to avoid bias by adding selected papers published after that date, Importantly, the message of these studies would not have changed the recommendations of the working group. As automated seizure detection using wearable devices is a rapidly developing field, we suggest updating this guideline at regular intervals or when high-level evidence is published, that would influence the recommendations.

Weighing the evidence

One comment addressed the way we weighed the published evidence and suggested that the recommendations were only based on the performance of the devices, and not on their clinical utility.

The way the recommendations were developed and the wording followed strictly the GRADE method. We found high-level of evidence that the validated wearable devices accurately detected tonic-clonic seizures (are “effective”). Conversely, there were only two “in-field studies” (phase 4) addressing its applicability, feasibility in the home environment, and assessing the clinical benefit of the devices. Although evidence is scarce, data from a phase-4 study suggest that a device decreased the number of injuries associated with tonic-clonic seizures.

In addition, based on the existing evidence regarding the association between nocturnal supervision and the risk of SUDEP, the practice guidelines of the American Academy of Neurology provide the following recommendation: “Recommendation 4. For persons with frequent GTCS and nocturnal seizures, clinicians may advise selected patients and families, if permitted by their individualized epilepsy and psychosocial circumstances, to use nocturnal supervision or other nocturnal precautions, such as the use of a remote listening device, to reduce SUDEP risk (Level C).” (Harden et al. Neurology 2017). We believe that GTCS detecting devices triggering an alarm

can be assimilated to remote listening devices, and that the above level C recommendation shall apply to both types of devices.

Altogether, we have high-level evidence that devices are “effective” in detecting tonic-clonic seizures, and we have scarce / indirect evidence about their clinical benefit. Following the GRADE methodology, this resulted in a weak, conditional recommendation. The wording of the recommendation reflects all the aspects listed above. We recommended using clinically validated wearable devices for automated detection of tonic-clonic seizures, especially in unsupervised patients who do not share a bedroom but where alarms can result in rapid intervention, within few minutes.

Wording of the recommendations

By unsupervised patients, we meant patients sleeping alone and without other form of supervision (for example CCTV). This was based on the risk-assessment: patients not sharing a bedroom and having at least one GTCS or FBTCS per year had a 67-fold increased risk of SUDEP (Sveinsson et al., Neurology 2020).

The wording of the recommendation is according to the GRADE method: the assessed intervention is either recommended or it is not recommended. For applications, lacking sufficient evidence at the time of the guideline development, the wording “does not recommend” should not hamper further research and development in the field. We revised the wording of the second recommendation to emphasize this, and explained it in more detail in the Discussion section of the revised text.

It was mentioned in the comments that in some patients, seizure types other than GTCS or FBTCS can be reliably detected using wearable devices. Indeed, we found studies supporting this statement (Table 2). However, these are only phase 2 studies, and there was lack of evidence for the feasibility of pre-selecting the suitable patients and for the clinical benefit. Therefore, the working group considered that these were not sufficient for issuing a positive recommendation. Further research and development are needed.

Concerning the granularity of the recommendations: The decision to recommend or not wearable devices is up to the physician managing the patients and this decision should be tailored to each individual patient. Therefore, these recommendations are meant to lay out guiding principles, and they allow some degree of flexibility.

Seizure quantification

Several comments noted that we focused the recommendations on automated seizure detection for triggering seizure-alarms, and we did not include recommendations for using automated seizure detection for seizure quantification. Actually, we specified in the Introduction section that the goals of the working group included reviewing the published evidence for using wearable devices to improve the objective documentation of seizure frequency. However, the current rate of false alarms might overestimate the true seizure frequency. Changes over time in FAR could lead to misleading within-patient trends of the detected event rate. We did not identify high-level published evidence for the accuracy or for the clinical benefit of seizure quantification using the currently available wearable devices. Therefore, based on the methodology we used, we were not able to issue a recommendation for this application – at present. However, we fully agree on the importance of objective seizure quantification and we listed this under "need for further research and development". Several approaches seem to be promising for solving the issue of false alarms for seizure quantification. Off-line visual analysis by experts, of the surface EMG signals automatically detected by an algorithm resulted in accurate validation of the events. Off-line analysis of the biosignals, using cloud-computing and artificial intelligence could provide more accurate seizure detection. We added this explanation to the Discussion section.

Need for more phase 4 studies

Several comments mentioned the importance of in-field studies. We fully agree with these comments and we listed this under the "need for further research and development". In addition, we added a paragraph to the Discussion section specifying that there is a need for more in-field studies for several reasons: 1) a more realistic estimation of the false alarm rate in the home environment of the patients; 2) assessment of the feasibility of ultra-long-term use of these devices, including patient groups with additional challenges (co-morbidities, disabilities); 3) estimation of the proportion of time with device deficiency; 4) investigation of the ultimate clinical benefit of wearing the devices; 5) adherence to daily usage. We agree that wearable devices will be of benefit only to the degree that patients accept their long-term and everyday use. Assessment of the device retention rate in phase 4 studies will be important.

Importance of high quality clinical studies

We fully agree with the comments that emphasized the need for high quality clinical studies. The crucial aspects mentioned in the comments are all included into the standards we have proposed

for clinical validation studies of wearable devices for automated seizure detection (Beniczky and Ryvlin, Epilepsia 2018).

Seizure detection vs. seizure prediction

Based on information from scalp EEG and video recording, the precise time-point of seizure-start often difficult to determine. Intracranial recordings demonstrated unequivocal seizure activity before any signs visible on the scalp EEG or video. Therefore, seizure-detection time-points shortly preceding the non-invasive video-EEG data, cannot be considered as seizure prediction. We would like to emphasize that seizure prediction was beyond the scope and limitation of this work.

Sándor Beniczky, Samuel Wiebe, Jesper Jeppesen, William O. Tatum, Milan Brazdil, Yuping Wang, Susan T. Herman and Philippe Ryvlin

References

Arends J, Thijs RD, Gutter T, Ungureanu C, Cluitmans P, Van Dijk J, et al. Multimodal nocturnal seizure detection in a residential care setting A long-term prospective trial. Neurology. 2018; 91(21):E2010–9.Beniczky and Ryvlin, Epilepsia 2018

Beniczky S, Ryvlin P. Standards for testing and clinical validation of seizure detection devices. Epilepsia. 2018;59 Suppl 1:9-13. doi: 10.1111/epi.14049.

Halford JJ, Sperling MR, Nair DR, Dlugos DJ, Tatum WO, Harvey J, et al. Detection of generalized tonic–clonic seizures using surface electromyographic monitoring. Epilepsia. 2017; 58(11):1861–9.Harden et al. Neurology 2017

Harden C, Tomson T, Gloss D, Buchhalter J, Cross JH, Donner E, et al. Practice guideline summary: Sudden unexpected death in epilepsy incidence rates and risk factors: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology. 2017; 88(17):1674–80.

Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav. 2011 Apr;20(4):638-41. doi: 10.1016/j.yebeh.2011.01.019.

Onorati F, Regalia G, Caborni C, Migliorini M, Bender D, Poh MZ, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia. 2017; 58(11):1870–9.

Regalia G, Onorati F, Lai M, Caborni C, Picard RW. Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands. Epilepsy Res. 2019;153:79-82. doi: 10.1016/j.eplepsyres.2019.02.007.

Sveinsson O, Andersson T, Mattsson P, Carlsson S, Tomson T. Clinical risk factors in SUDEP: A nationwide population-based case-control study. Neurology. 2020; 94(4):e419–29.

van Westrhenen A, Petkov G, Kalitzin SN, Lazeron RHC, Thijs RD. Automated video-based detection of nocturnal motor seizures in children. Epilepsia. 2020; doi: 10.1111/epi.16504.

Vandecasteele K, De Cooman T, Dan J, Cleeren E, Van Huffel S, Hunyadi B, Van Paesschen W. Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels. Epilepsia. 2020;61(4):766-775.

Weisdorf S, Duun-Henriksen J, Kjeldsen MJ, Poulsen FR, Gangstad SW, Kjaer TW. Ultra-long-term subcutaneous home monitoring of epilepsy-490 days of EEG from nine patients. Epilepsia. 2019 Nov;60(11):2204-2214. doi: 10.1111/epi.16360.