HHS Public Access

Author manuscript

Epilepsia. Author manuscript; available in PMC 2025 April 01.

Published in final edited form as:

Epilepsia. 2024 April ; 65(4): 1017–1028. doi:10.1111/epi.17917.

Minimum clinical utility standards for wearable seizure detectors – a simulation study

Daniel M. Goldenholz, MD, PhD[1,2], **Philippa J. Karoly, PhD**[3], **Pedro F. Viana, MD**[4,5], **Ewan Nurse, PhD**[6], **Tobias Loddenkemper, MD**[1,7], **Andreas Schulze-Bonhage, MD**[8], **Solveig Vieluf, PhD**[1,7], **Elisa Bruno, MD, PhD**[4], **Mona Nasseri, PhD**[9], **Mark P. Richardson, FRCP PhD**[4,*], **Benjamin H. Brinkmann, PhD**[9,*], **M. Brandon Westover, MD, PhD**[1,2,10,11,*]

[1] –Department of Neurology, Harvard Medical School, Boston USA

[2] –Department of Neurology, Beth Israel Deaconess Medical Center, Boston USA

[3] –Department of Neurology, University of Melbourne

[4] -School of Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom

[5] –Faculty of Medicine, University of Lisbon, Portugal.

[6] –Seer Medical, Australia

[7] –Department of Neurology, Boston Children's Hospital, Boston USA

[8] –Epilepsy Center, University Medical Center Freiburg-University of Freiburg, Germany,

[9] -Department of Neurology, Mayo Clinic, Rochester MN USA

[10]-Department of Neurology, Massachusetts General Hospital, Boston USA

[11]-McCace Center, Boston USA

Abstract

daniel.goldenholz@bidmc.harvard.edu .

*senior co-authors

Author contributions

DMG – project design, software development, data analysis, data interpretation figures, writing. DMG has direct access and verified the underlying data reported.
PJK – project design, data interpretation, editing manuscript.
PFV – literature search, data interpretation, editing manuscript.
EN – data interpretation, editing manuscript.
TL – project design, literature search, data interpretation, editing manuscript.
ASB – data interpretation, editing manuscript.
SV – data interpretation, editing manuscript.
EB – data interpretation, editing manuscript.
MN – data interpretation, editing manuscript.
MPR – project oversight, project design, data interpretation, editing manuscript.
BHB – project oversight, project design, data interpretation, editing manuscript.
MBW – project oversight, project design, data interpretation, editing manuscript.

Ethical publication statement:

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Objective: Epilepsy management employs self-reported seizure diaries, despite evidence of seizure under-reporting. Wearable and implantable seizure detection devices are now becoming more widely available. There are no clear guidelines about what levels of accuracy are sufficient. This study aimed to simulate clinical use cases and identify the necessary level of accuracy for each.

Methods: Using a realistic seizure simulator (CHOCOLATES) a ground truth was produced, which was then sampled to generate signals from simulated seizure detectors of various capabilities. Five use cases were evaluated: (1) randomized clinical trials (RCT), (2) medication adjustment in clinic, (3) injury prevention, (4) sudden unexpected death in epilepsy (SUDEP) prevention, and (5) treatment of seizure clusters. We considered sensitivity (0–100%), false alarm rate (FAR) (0–2/day), and device type (external wearable vs. implant) in each scenario.

Results: The RCT case was efficient for a wide range of wearable parameters, though implantable devices were preferred. Lower accuracy wearables resulted in subtle changes in the distribution of patients enrolled in RCTs, and therefore higher sensitivity and lower FAR values were preferred. In the clinic case, a wide range of sensitivity, FAR and device type yielded similar results. For injury prevention, SUDEP prevention and seizure cluster treatment, each scenario required high sensitivity and yet was minimally influenced by FAR.

Significance: The choice of use case is paramount in determining acceptable accuracy levels for a wearable seizure detection device. We offer simulation results for determining and verifying utility for specific use case and specific wearable parameters.

INTRODUCTION

Epilepsy afflicts over 46 million people worldwide[1], yet roughly one in three suffer from ongoing seizures despite treatment[2]. Even though the disease is defined by the presence of seizures, self-reported seizure records may undercount events by 50%[3]. Thus, having objective methods of quantifying disease burden has been a longstanding goal that is beginning to see new developments in recent years. Wearable and implantable devices now document seizures using biosignals including: electrodermal activity[4], accelerometers[5], electromyography[6] and subscalp EEG[7]. We consider devices capable of detecting seizures using a non-invasive wearable or an implanted instrument collectively as “wearables” for the purpose of this study. While some studies discuss the use of wearable in clinical trials[8,9], there are no clear details on what degree of accuracy is acceptable and in which clinical context. Other standards for clinical validation and testing of seizure detection wearables[10] similarly do not provide needed guidance. Regulatory approval of medical devices for seizure detection appears to be focused on the evidence of a high detection performance[11]. As wearable technology becomes more affordable and accessible, there is an urgent need to determine under what degree of accuracy and in which settings these tools should be deployed.

This study aimed to address this need with large-scale, realistic simulations of seizure diaries. To facilitate a more informed discussion, we considered five practical use cases for wearable seizure detectors: (1) randomized clinical trials (RCTs), (2) medication adjustment in clinic, (3) injury prevention, (4) sudden unexpected death in epilepsy

(SUDEP) prevention, and (5) seizure cluster treatment. We used a new realistic seizure diary simulator (CHOCOLATES) that accounts for multiple statistical features seen in seizure diaries[12]. We then simulated each use case to better define the characteristics needed for each specific clinical application. The advantage of simulation rather than a series of formal clinical trials is that it allows for millions of synthetic patients to be exposed to virtual risks without harming anyone, while building important expectations about how wearables can be effectively used in real patients.

METHODS

All cases

In all use cases below, the simulations were based on CHOCOLATES[12] (Appendix A). CHOCOLATES is an open-source simulator capable of generating seizure diaries that have five characteristics observed in numerous clinical studies. These features are: (1) heterogeneity in average seizure rates across subjects[13], (2) a consistent relationship between average and standard deviation in seizure rate across patients (the "L-relationship")[14], (3) multiple coexisting seizure risk cycles[15,16], (4) seizure clustering features[17,18], and (5) limitations on minimum inter-seizure intervals[13,19,20]. The output of CHOCOLATES is a series of seizure counts, representing the number of seizures the synthetic patient reports within a user-specified time window (such as daily, or hourly). CHOCOLATES is able to reproduce typical self-reported seizure diaries from a heterogeneous community of patients. However, self-reported seizure diaries can suffer from under-reporting, a finding reproduced across many studies with different modalities[3]. Therefore, we developed an evidence-informed pipeline (Appendix B) using CHOCOLATES (Appendix A). The pipeline produced several distinct kinds of diaries: true electrographic and clinical seizures ("true e-szs" and "true clin szs"), wearable detected electrographic and clinical seizures ("detected e-sz" and "detected clinical sz"), and clinically self-reported seizures ("observed sz"). Self-report was assumed to have a sensitivity of 0.5 (based on under-reporting[3]) and a false alarm rate of 0 (based on lack of detailed data, see Appendix B). This overall approach allows for the comparison of seizure reports from various sources when a ground truth is known. We tested a variety of sensitivity values from 0.5 up to 1, including false alarm rates (FAR) from 0 to 6 alarms/day, based on currently existing wearable device studies[21–28] (Appendix B, Table S1).

RCT use case

The RCT case approximates a phase II or III trial testing the efficacy of a new therapy for epilepsy in a randomized placebo controlled parallel design, using 2 months of baseline and 3 months of steady-state testing[29]. Clinical trials were simulated as in previously published reports[30–36]. RCTs followed parallel placebo-controlled design, consistent with modern seizure therapy trials[29,37]. Trials included 50% placebo patients and 50% drug patients. We assumed a 20% efficacy for all drugs across the population[38], meaning that on average, drug has a 20% probability of preventing any individual seizure (Appendix C). During the 2-month baseline period eligible patients needed at least 4 seizures/month (according to the wearable). A correction factor ensured removal of the average expected number of false alarms from trial data before computing trial success. Trial metrics included

50%-responder rate (RR50) and median percentage change (MPC)[39]. Trial success indicated the ability to differentiate drug from placebo ($p<0.05$). The average success rate in 10,000 simulated RCTs represented the statistical power. The simulations were repeated to identify the minimum number of patients (N) needed to obtain 90% power for a given wearable device and a given trial metric. We compared all values to the minimum N value estimated for an RCT with self-reported seizures (assuming FAR=0, sensitivity=0.5)[3]. All reported values were percentages of the self-reported minimum N. Thus, if a wearable produced a minimum N that is equal to that of self-reported seizures, the result reported (Figure 1) was 100. Further methodological details are in Appendix C.

Medication adjustment in clinic use case

The clinic case reflected the situation of a patient who sees the physician every 3 months and changes medication if appropriate. The choice to increase or decrease medications was based on a standardized algorithm agreed upon by consensus of the authors. The algorithm represented an approximation to clinical practice.

Let the current and prior 3-month seizure count be $C_{CURRENT}$ and C_{PRIOR} and $P_{increase}$ be the probability of choosing to increase medication:

1. If $C_{CURRENT} >$ (½) C_{PRIOR} then increase medication with probability $P_{increase}$.
2. If $C_{CURRENT} = 0$ or $C_{CURRENT} <=$ (½) C_{PRIOR} then no change.
3. If no medication change has been made for 2 years (i.e., consistently doing well according to rule (2)), then decrease medication.
4. Never decrease from rule (3) to no drugs.

All medications were assumed to either be full dose or half dose. This does not perfectly match up with some anti-seizure medications (ASMs) that have multiple dose options, however it roughly aligns with most. This simplification is reasonable because almost all ASMs at maximal dose are roughly 20% more effective than placebo[38], and because distinguishing improvement smaller than 10% is statistically challenging.

Medication increases depend on the patient's current regimen. If a patient is on a half dose of a medication, an increase results in a dose increase to full dose. Conversely, if they are already on a full dose, it results in the addition of another concurrent medication at half the maximal dose. In the same way, medication decreases for a drug at full dose would decrease to half dose and decreases on half doses would remove that medication. It is important to note that case (3) of medication decrease only would occur in the situation of a patient who was found to be either seizure free or decreasing in their seizure rate for the entire 2-year period. It is worth noting that the model presented here assumed medications are provided concurrently, rather than swapping one medication for another. It is unknown if either strategy is optimal[40].

The simulation accounted for the possibility of new medications resulting in seizure freedom[2] which is either lasting, fleeting, or fluctuating[41]. Based on prior research and typical clinical practice, no patient was allowed more than 6 medications at once[2]. All medications were modelled to be either 10% (at half dose) or 20% (at full dose) effective

at reducing seizures. As with the RCT case, a correction for expected false alarms was included when considering change in seizure rates at each clinic visit. A total of 10,000 patients were simulated daily for 10 years for each wearable evaluated. Outcomes were the average monthly seizures per patient and average monthly drug exposure per patient over the entire 10 years. Further details are found in Appendix D.

Injury prevention use case

To simulate the value of a wearable detecting a seizure fast enough to prevent injury, we assumed that the detection time was extremely fast, providing sufficient warning to the patient or caregiver to prevent an injury. We used a memoryless random variable with a low or high injury rate, either 4.8 or 296 per 100 patient-years[42,43], to select times for injury. When the injury variable became positive, the injury occurred at the time of the next seizure from the true clinical seizure diary. The wearable event detection at that seizure determined if the injury was detected. As a simplification, we assumed that if the wearable detected a seizure destined to injure, that detection would always prevent that injury. A set of 10,000 synthetic patients simulated over 10 years (10 minutes per sample) was generated for each wearable tested. The outcome was the average injury rate per patient over the 10-year simulated timeframe. Further details are provided in Appendix E.

SUDEP prevention use case

The SUDEP case evaluated the possibility of detecting a generalized tonic clonic seizure (GTC) that was associated with SUDEP using a wearable. We assumed that the wearable could alert a caregiver to rapidly provide rescue maneuvers to prevent the SUDEP if the GTC was detected. GTCs are known to occur in only 23% of people with epilepsy[44], and only then 9–18% of the time[45]. The SUDEP base rate[46] is 1.2 in 1000 patient-years, which increases with GTCs occurring within the most recent 12 months. If there are 1–2 yearly GTCs, the rate becomes 6.1 in 1000; if there are 3 or more GTCs, the rate becomes 18 in 1000. Using these published risk rates based on GTCs of each individual diary and a memoryless random variable to generate SUDEP times, CHOCOLATES generated 100,000 synthetic patients with 10 years of diary each. When a possible SUDEP was simulated, it was moved to the nearest next true seizure (of any type), and detection of that seizure was tested based on the specific wearable. Rates of detected and missed SUDEP for different wearable configurations were tallied. Further details are found in Appendix F.

Seizure cluster treatment use case

The cluster treatment case evaluated the value of using a wearable to determine if a rescue medication should be administered during a seizure cluster. Clusters consisted of 2 or more seizures within a 6-hour period[47]. If the wearable detected a cluster, patients received a rescue medication with 20–30% efficacy that persisted for 6–24 hours[47–50]. CHOCOLATES generated diary samples every 10 minutes for 10 years including each of 10,000 patients in each wearable configuration. Outcomes were the number of seizures prevented and number of rescue medications taken. Further detail is available in Appendix G.

DATA AVAILABILITY

Add data used in this study was synthetically generated using the open source code.

CODE AVAILABILITY

The underlying code for this study is available on GitHub and can be accessed via this link https://github.com/GoldenholzLab/WEARsimulator/.

RESULTS

The RCT use case (Figure 1) demonstrated that a large range of sensitivity and FAR values produced RCT efficiencies that were better than or comparable to self-report (exact numbers are given in Appendix C). There was a modest effect of sensitivity, i.e., despite a large range of sensitivity, the number of patients needed does not change much. This is because the RCTs have an eligibility determination based on the wearable output. The lower trial power seen with lower sensitivity wearables is partially offset by demanding a concomitant higher true baseline rate from eligible participants. This effect is discussed in more detail in Appendix C. The clinical+electrographic (implant) case resulted in higher trial efficiency for any accuracy configuration. The intra-subject variability for FAR was tested over a range of values, and it was noted that as this variability increased, RCT efficiency decreased (Appendix H).

The clinic case simulations (Figure 2, and Appendix D) found that higher sensitivity resulted in higher average number of drugs taken, and smaller average true seizure rates. With increasing FAR values, average number of drugs increased, and average true seizure rates increased. Basing clinical decisions on wearable detections resulted in a higher average drug burden compared to basing clinical decisions on self-reported events, although with a reduced average seizure frequency provided the detectors FAR was 1 alarm/day or lower.

The injury prevention simulations found that regardless of accident rate or wearable type (implanted or external wearable), the detection rate very closely matched the sensitivity (Table 1, and Appendix E), as expected. High sensitivity, therefore, produced lower injury rates.

The SUDEP prevention simulations showed (Table 2 and Appendix F) that regardless of seizure type (clinical or clinical+electrographic), the prevented fraction of SUDEP events nearly matched the sensitivity of the wearable, as expected. It is notable that rarely the prevented fraction slightly exceeds the sensitivity of the device by 1–2% - this likely represents statistical noise in the simulation.

The cluster treatment use case results (Figure 3) show sensitivity to be the most important factor in determining average rescue drugs used as well as average rescued seizures, whereas the false alarm rate had minimal effect on both.

DISCUSSION

The present study explored several different use cases for wearable seizure detection via realistic seizure diary simulation. Overall, the findings demonstrate the utility of wearable seizure detectors within a plausible range of sensitivity and FAR criteria, particularly for RCTs and clinical management. Sensitivity and FAR results have been published for several different seizure monitoring systems, and our simulations were designed to cover the reported ranges in these data (see Supplementary Table S1) as well as to explore the limits of acceptability of different sensitivity and FAR rates in different contexts. This study highlights the importance of selecting a wearable based on the desired use case, rather than requiring ideal accuracy.

Some of the present results could be predicted with simple statistics, such as the injury case. Others would be challenging to anticipate because of the numerous factors involved. The present study highlights the value of simulating extremely large sets of conditions without exposing patients to any risk. Clinically realistic values for some seizure detector accuracy values are presented in the Appendix, which may serve as a guide of some of the currently clinically available tools. We covered this range of sensitivity/FAR and beyond in our simulations to provide a broad overview of what can be expected.

The RCT case (Figure 1) found that a wide range of sensitivity and FAR values could be acceptable, with a caveat that increased FAR or decreased sensitivity may result in more challenging enrollment due to effectively higher eligibility criteria to compensate for device inaccuracy (Appendix C). Assuming the right patients would be found, the power of the trial would still be relatively acceptable. Implanted devices resulted in higher trial efficiency, as expected[35], though it is unknown if treatments capable of reducing electrographic seizures would have analogous results with clinical seizures[51]. It is important to note that traditional RCTs (based entirely on self-reported diaries) include human validation and reconciliation. Thus, wearable-based RCTs would need to either conduct an analogous data-validity check based entirely on wearable data, or some kind of hybrid approach that includes self-report. Our analysis assumes a purely wearable data-driven approach.

The clinic use case (Figure 2) was perhaps the most forgiving; wearables with a wide range of sensitivity and FAR values proved similar to self-reports without caveat. The injury prevention, SUDEP prevention and cluster treatment cases each demonstrated that high sensitivity, regardless of FAR, is of critical importance. However, user tolerance of higher FAR would likely place a ceiling on acceptable FAR, irrespective of its importance in these simulations. In other words, for any use case, if a FAR is too high, the device that called wolf would be the primary problem.

This study also highlights the critical importance of publishing (and regularly updating) the FAR performance for wearable devices as obtained from validation studies. Both the RCT and clinic simulations showed dramatic improvement by correcting for expected false alarms, since it was assumed the trialist or clinician would be aware of the published FAR value of the specific wearable. This correction allows for devices with seemingly large FAR values to still be valuable in the appropriate context. While we were not able to incorporate

this factor in our simulations due to the limited data available on adherence rates and FAR, it is an important practical consideration in all use cases, although in the RCT case if no patient action is required these effects may be minimal.

Furthermore, the presented simulations revealed that there is value in characterizing intra-subject FAR variability for a wearable, as RCTs efficiency decreases with greater FAR variability (Appendix H). This is not a value typically reported currently; however, we recommend both inter-subject and intra-subject FAR values to be reported to better understand the impact of wearables for specific use cases.

This study had several limitations. The foundation of our simulations is the CHOCOLATES software. CHOCOLATES represents state-of-the-art simulation tools, because it incorporates more of what is known statistically about seizure diaries than any prior tool[12]. However, it is a very complex model, and may have unknown deficiencies. In order to undertake a large-scale simulation such as this several assumptions are necessary, and it is impossible to accurately account for every possible parameter in every situation. Wherever possible we have relied on published data to guide our choices of input parameters and have simulated values across a broad range of realistic possibilities. For example, the clinical treatment algorithm considered in our study represents merely one candidate strategy for treating epilepsy clinically, and without formal guidelines many clinicians develop their own personal strategies. Our assumption the self-reported seizures represents a sensitivity of 0.5 and FAR of 0 in particular may be based on incomplete information. The literature on reporting errors in patient diaries has mainly focused on under-reporting events, although some studies[52,53] suggest false positive diary reports can occur, and even in the context of seizure under-reporting on average[3,54,55] some false positive reports are possible. However, no clear data exists in the literature to guide the choice of a false alarm rate for diaries and given the net underreporting of seizures we have used a net FAR of zero in our diary case. By choosing to simulate with FAR=0, we are not asserting this as a “true” value for FAR, rather we are simply assuming this value in the absence of more definitive data. Furthermore, some specific patients may have highly individualized FAR values. For example, daily joggers versus sedentary patients are likely to differ in their daytime FAR for wearables. Our simulations did not account for individualized FAR values, but this may become important for certain devices. Similarly, our study assumed perfect adherence and signal quality of each device, however, realistic values for these are seldom reported[9]. Another factor to consider about FAR is patients experiencing frequent false alarms with a wearable would be more likely to discontinue use. Therefore, the effective FAR value for a given device may be artificially low if the denominator ignores such drop-out. Our analysis can be considered in the context of a true FAR value, and adjusted for whatever dropout rate is expected in realistic populations. We acknowledge that in the SUDEP and injury cases, the assumption that every detected seizure would result in intervention quickly enough may be too optimistic and would require control over additional factors. Moreover, SUDEP risk may be elevated at night[56] but our simulation did not account for day/night differences. Accounting for nocturnal only events in a model that equally produces day or night seizures (on average) would not be expected to change the outcome. Importantly, the simulations (other than SUDEP) did not distinguish between seizure subtypes, an area that has been incompletely studied and currently only partial information is available about

relative prevalence of different seizure subtypes. The SUDEP simulations also assumed a perfect rescue rate after seizure detection, which unfortunately is overly optimistic. It is known that SUDEP can occur even in highly observed medical settings[57]. Our purpose here was to assume a “best case” scenario of perfect rescue, which could then be taken as an upper limit of expected value for SUDEP prevention with wearables.

The majority of data about seizure patterns were obtained primarily from patients with drug-resistant epilepsy. Therefore, conclusions based on the present simulations would have analogies but not direct comparisons for patients who appear to be drug-sensitive. All these concerns may be overcome in the future. Our simulations are open-source, and as new information becomes available, these results can be updated by interested investigators. Finally, personal preference of the users’ group (e.g. people with epilepsy, caregivers) and of the single individual is another factor to consider as it could additionally influence the criteria for acceptable performance in the real world setting[11].

In conclusion, we present a realistic large-scale set of simulations to evaluate the relative merits of different configurations of wearable and implantable seizure detectors, depending on the use case. We found case-specific differences in desirable properties, which we hope will help guide device designers, regulators, and clinicians.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding was from NIH (K23NS124656, KL2TR002542, UG3NS123066, U44NS121562, R01NS102190, R01NS102574, R01NS107291, RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598) NSF 2014431, CBET-2138378, Epilepsy Foundation of America, American Epilepsy Society, National Health & Medical Research Council GNT1178220. In all cases, the funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Disclosures of potential conflicts of interest:

DMG is an unpaid advisor for Epilepsy AI and Eysz. He has been a paid advisor for Magic Leap. He has been provided speaker fees from AAN, AES and ACNS. He also previously has been a paid consultant for Neuro Event Labs, IDR, LivaNova and Health Advances.

PJK, ESN has a financial interest in Seer Medical Pty. Ltd.

TL is part of patents and patent applications to detect and predict clinical outcomes, and to detect, manage, diagnose and treat neurological conditions, epilepsy, and seizures. He received device donations from various companies, including Empatica. He declares no non-financial competing interests.

SV is part of a patent application covering technology for seizure forecasting, but declares no non-financial competing interests.

PFV has received travel fees and payment for an unrelated research task from UNEEG Medical A/S, but declares no non-financial competing interests.

EB has no conflicts.

BHB has received research support from UNEEG Medical A/S and Seer Medical Pty Ltd, and has received device donations for research from Medtronic Inc. He has licensed IP to Cadence Neurosciences, and has consulted for Otsuka Pharmaceuticals, but declares no non-financial competing interests.

MPR has consulted for UNEEG Medical A/S. He has licensed IP to Neuronostics Ltd but declares no non-financial competing interests.

MN has no conflicts.

MBW is a co-founder of Beacon Biosignals but declares no non-financial competing interests.

All other authors declare no financial or non-financial competing interests.

REFERENCES

1. Beghi E The Epidemiology of Epilepsy. Neuroepidemiology [Internet]. 2020 [cited 2022]; 54(2):185–91. Available from: https://www.karger.com/Article/FullText/503831
2. Chen Z, Brodie MJ, Liew D, Kwan P. Treatment outcomes in patients with newly diagnosed epilepsy treated with established and new antiepileptic drugs a 30-year longitudinal cohort study. JAMA Neurol. 2018; 75(3):279–86. [PubMed: 29279892]
3. Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol [Internet]. 2018; 17(3):279–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29452687 [PubMed: 29452687]
4. 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. [PubMed: 28980315]
5. Beniczky S, Polster T, Kjaer TW, Hjalgrim H. Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: A prospective, multicenter study. Epilepsia. 2013; 54(4):58–61. [PubMed: 22708895]
6. Beniczky S, Conradsen I, Henning O, Fabricius M, Wolf P. Automated real-time detection of tonic-clonic seizures using a wearable EMG device. Neurology. 2018; 90(5):e428–34. [PubMed: 29305441]
7. Duun-Henriksen J, Baud M, Richardson MP, Cook M, Kouvas G, Heasman JM, et al. A new era in electroencephalographic monitoring? Subscalp devices for ultra-long-term recordings. Epilepsia. 2020; 61(9):1805–17. [PubMed: 32852091]
8. Kurada AV, Srinivasan T, Hammond S, Ulate-Campos A, Bidwell J. Seizure detection devices for use in antiseizure medication clinical trials: A systematic review. Seizure [Internet]. 2019 [cited 2023]; 66:61–9. Available from: https://pubmed.ncbi.nlm.nih.gov/30802844/
9. Beniczky S, Wiebe S, Jeppesen J, Tatum WO, Brazdil M, Wang Y, et al. Automated seizure detection using wearable devices: A clinical practice guideline of the International League Against Epilepsy and the International Federation of Clinical Neurophysiology. Clin Neurophysiol [Internet]. 2021 [cited 2023]; 132(5):1173–84. Available from: https://pubmed.ncbi.nlm.nih.gov/33678577/
10. Beniczky S, Ryvlin P. Standards for testing and clinical validation of seizure detection devices. Epilepsia [Internet]. 2018 [cited 2023]; 59 Suppl 1:9–13. Available from: https://pubmed.ncbi.nlm.nih.gov/29873827/
11. Bruno E, Viana PF, Sperling MR, Richardson MP. Seizure detection at home: Do devices on the market match the needs of people living with epilepsy and their caregivers? Epilepsia [Internet]. 2020 [cited 2023]; 61 Suppl 1(S1):S11–24. Available from: https://pubmed.ncbi.nlm.nih.gov/32385909/
12. Goldenholz DM, Westover MB. Flexible realistic simulation of seizure occurrence recapitulating statistical properties of seizure diaries. Epilepsia [Internet]. 2023; 64(2):396–405. Available from: http://www.ncbi.nlm.nih.gov/pubmed/36401798 [PubMed: 36401798]
13. Ferastraoaru V, Goldenholz DM, Chiang S, Moss R, Theodore WH, Haut SR. Characteristics of large patient-reported outcomes: Where can one million seizures get us? Epilepsia Open [Internet]. 2018; 3(3):364–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30187007 [PubMed: 30187007]
14. Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, et al. Is seizure frequency variance a predictable quantity? Ann Clin Transl Neurol. 2018; 5(2):201–7. [PubMed: 29468180]

15. Baud MO, Kleen JK, Mirro EA, Andrechak JC, King-Stephens D, Chang EF, et al. Multi-day rhythms modulate seizure risk in epilepsy. Nat Commun [Internet]. 2018; 9(1):1–10. Available from: 10.1038/s41467-017-02577-y [PubMed: 29317637]
16. Karoly PJ, Goldenholz DM, Freestone DR, Moss RE, Grayden DB, Theodore WH, et al. Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. Lancet Neurol [Internet]. 2018; 17(11):977–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30219655 [PubMed: 30219655]
17. Chiang S, Haut SR, Ferastraoaru V, Rao VR, Baud MO, Theodore WH, et al. Individualizing the definition of seizure clusters based on temporal clustering analysis. Epilepsy Res. 2020; 163.
18. Haut SR. Seizure clusters: characteristics and treatment. Curr Opin Neurol [Internet]. 2015; 28(2):143–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25695133 [PubMed: 25695133]
19. Trinka E, Cock H, Hesdorffer D, Rossetti AO, Scheffer IE, Shinnar S, et al. A definition and classification of status epilepticus - Report of the ILAE Task Force on Classification of Status Epilepticus. Epilepsia. 2015; 56(10):1515–23. [PubMed: 26336950]
20. Theodore WH, Porter RJ, Albert P, Kelley K, Bromfield E, Devinsky O, et al. The secondarily generalized tonic-clonic seizure: a videotape analysis. Neurology [Internet]. 1994; 44(8):1403–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8058138 [PubMed: 8058138]
21. 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. [PubMed: 30846346]
22. 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 [Internet]. 2017; 58(11):1861–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28980702 [PubMed: 28980702]
23. Jeppesen J, Fuglsang-Frederiksen A, Johansen P, Christensen J, Wüstenhagen S, Tankisi H, et al. Seizure detection based on heart rate variability using a wearable electrocardiography device. Epilepsia. 2019; 60(10):2105–13. [PubMed: 31538347]
24. 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. [PubMed: 30355702]
25. Baumgartner C, Koren JP. Seizure detection using scalp-EEG. Epilepsia. 2018; 59 Suppl 1:14–22. [PubMed: 29873826]
26. Narechania AP, Garić II, Sen-Gupta I, Macken MP, Gerard EE, Schuele SU. Assessment of a quasi-piezoelectric mattress monitor as a detection system for generalized convulsions. Epilepsy Behav. 2013; 28(2):172–6. [PubMed: 23747502]
27. Arends JB, van Dorp J, van Hoek D, Kramer N, van Mierlo P, van der Vorst D, et al. Diagnostic accuracy of audio-based seizure detection in patients with severe epilepsy and an intellectual disability. Epilepsy and Behavior [Internet]. 2016; 62:180–5. Available from: 10.1016/j.yebeh.2016.06.008 [PubMed: 27490905]
28. Nasseri M, Pal Attia T, Joseph B, Gregg NM, Nurse ES, Viana PF, et al. Non-invasive wearable seizure detection using long-short-term memory networks with transfer learning. J Neural Eng. 2021; 18(5).
29. Perucca E Antiepileptic drugs: evolution of our knowledge and changes in drug trials. Epileptic Disord [Internet]. 2019; 21(4):319–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31403463 [PubMed: 31403463]
30. Romero J, Goldenholz DM. Statistical efficiency of patient data in randomized clinical trials of epilepsy treatments. Epilepsia [Internet]. 2020 [cited 2020]; 61(8):1659–67. Available from: https://pubmed.ncbi.nlm.nih.gov/32658327/ [PubMed: 32658327]
31. Romero J, Chiang S, Goldenholz DM. Can machine learning improve randomized clinical trial analysis? Seizure [Internet]. 2021; 91:499–502. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34365104 [PubMed: 34365104]

32. Oliveira A, Romero JM, Goldenholz DM. Comparing the efficacy, exposure, and cost of clinical trial analysis methods. Epilepsia [Internet]. 2019; 60(12):e128–32. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.16384 [PubMed: 31724165]
33. Romero J, Larimer P, Chang B, Goldenholz SR, Goldenholz DM. Natural variability in seizure frequency: Implications for trials and placebo. Epilepsy Res [Internet]. 2020 [cited 2020]; 162:106306. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0920121119303341?via%3Dihub [PubMed: 32172145]
34. Goldenholz DM, Tharayil J, Moss R, Myers E, Theodore WH. Monte Carlo simulations of randomized clinical trials in epilepsy. Ann Clin Transl Neurol. 2017; 4(8):544–52. [PubMed: 28812044]
35. Goldenholz DM, Tharayil JJ, Kuzniecky R, Karoly P, Theodore WH, Cook MJ. Simulating clinical trials with and without intracranial EEG data. Epilepsia Open. 2017; 2(2):156–61. [PubMed: 28758158]
36. Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, et al. Does accounting for seizure frequency variability increase clinical trial power? Epilepsy Res [Internet]. 2017; 137(June):145–51. Available from: 10.1016/j.eplepsyres.2017.07.013 [PubMed: 28781216]
37. Perucca E What clinical trial designs have been used to test antiepileptic drugs and do we need to change them? Epileptic Disorders. 2012; 14(2):124–31. [PubMed: 22977898]
38. Rheims S, Cucherat M, Arzimanoglou A, Ryvlin P. Greater response to placebo in children than in adults: A systematic review and meta-analysis in drug-resistant partial epilepsy. PLoS Med. 2008; 5(8):1223–37.
39. Siddiqui O, Hershkowitz N. Primary Efficacy Endpoint in Clinical Trials of Antiepileptic Drugs: Change or Percentage Change. Drug Inf J [Internet]. 2010; 44(3):343–50. Available from: http://link.springer.com/10.1177/009286151004400316
40. Hakeem H, Alsfouk BAA, Kwan P, Brodie MJ, Chen Z. Should substitution monotherapy or combination therapy be used after failure of the first antiseizure medication? Observations from a 30-year cohort study. Epilepsia [Internet]. 2023 [cited 2023]; 64(5):1248–58. Available from: https://pubmed.ncbi.nlm.nih.gov/36869855/ [PubMed: 36869855]
41. Brodie MJ, Barry SJE, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. Neurology. 2012; 78(20):1548–54. [PubMed: 22573629]
42. Russell-Jones DL, Shorvon SD. The frequency and consequences of head injury in epileptic seizures. J Neurol Neurosurg Psychiatry. 1989; 52(5):659–62. [PubMed: 2732739]
43. Neufeld MY, Vishne T, Chistik V, Korczyn AD. Life-long history of injuries related to seizures. Epilepsy Res. 1999; 34(2–3):123–7. [PubMed: 10210026]
44. Hauser WA, Annegers JF, Rocca WA. Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota. Mayo Clin Proc. 1996; 71(6):576–86. [PubMed: 8642887]
45. Ferastraoaru V, Goldenholz DM, Chiang S, Moss R, Theodore WH, Haut SR. Characteristics of large patient-reported outcomes: Where can one million seizures get us? Epilepsia Open [Internet]. 2018; 3(3):364–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30187007 [PubMed: 30187007]
46. 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 [Internet]. 2017; 88(17):1674–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28438841 [PubMed: 28438841]
47. Spencer DC, Sinha SR, Choi EJ, Cleveland JM, King A, Meng T-C, et al. Safety and efficacy of midazolam nasal spray for the treatment of intermittent bouts of increased seizure activity in the epilepsy monitoring unit: A double-blind, randomized, placebo-controlled trial. Epilepsia. 2020; 61(11):2415–25. [PubMed: 33140403]
48. Detyniecki K, Van Ess PJ, Sequeira DJ, Wheless JW, Meng T-C, Pullman WE. Safety and efficacy of midazolam nasal spray in the outpatient treatment of patients with seizure clusters-a

randomized, double-blind, placebo-controlled trial. Epilepsia. 2019; 60(9):1797–808. [PubMed: 31140596]

49. FDA. Valtoco - FDA Prescribing information [Internet]. 2022 [cited 2022]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/211635s007lbl.pdf
50. FDA. Nayzilam - FDA PRESCRIBING INFORMATION [Internet]. 2019 [cited 2022]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/211321s000lbl.pdf
51. French JA. Improving clinical trial efficiency: Is technology the answer? Epilepsia Open [Internet]. 2017 [cited 2023]; 2(2):121–2. Available from: https://pubmed.ncbi.nlm.nih.gov/29588941/
52. Tatum WO, Winters L, Gieron M, Passaro EA, Benbadis S, Ferreira J, et al. Outpatient Seizure Identification. Journal of Clinical Neurophysiology. 2001; 18(1):14–9. [PubMed: 11290934]
53. DuBois JM, Boylan LS, Shiyko M, Barr WB, Devinsky O. Seizure prediction and recall. Epilepsy and Behavior [Internet]. 2010; 18(1–2):106–9. Available from: 10.1016/j.yebeh.2010.03.011 [PubMed: 20457544]
54. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study. Lancet Neurol. 2013; 12(6):563–71. [PubMed: 23642342]
55. Fisher RS, Blum DE, DiVentura B, Vannest J, Hixson JD, Moss R, et al. Seizure diaries for clinical research and practice: limitations and future prospects. Epilepsy Behav [Internet]. 2012; 24(3):304–10. Available from: 10.1016/j.yebeh.2012.04.128 [PubMed: 22652423]
56. Sveinsson O, Andersson T, Mattsson P, Carlsson S, Tomson T. Clinical risk factors in SUDEP: A nationwide population-based case-control study. Neurology [Internet]. 2020 [cited 2023]; 94(4):e419–29. Available from: https://pubmed.ncbi.nlm.nih.gov/31831600/
57. Ryvlin P, Nashef L, Lhatoo SD, Bateman LM, Bird J, Bleasel A, et al. Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): A retrospective study. Lancet Neurol. 2013; 12(10):966–77. [PubMed: 24012372]

KEY POINTS:

- Using realistic seizure diary simulation, we simulated different levels of accuracy for wearables and implanted seizure detection devices.
- Tested: (1) trials, (2) med adjustment in clinic, (3) injuries, (4) sudden unexpected death in epilepsy prevention, (5) meds for clusters
- Limits on how sensitive and how tolerant of false alarms a device are shown for each tested use
- This study can help guide utility of wearables depending on intended use

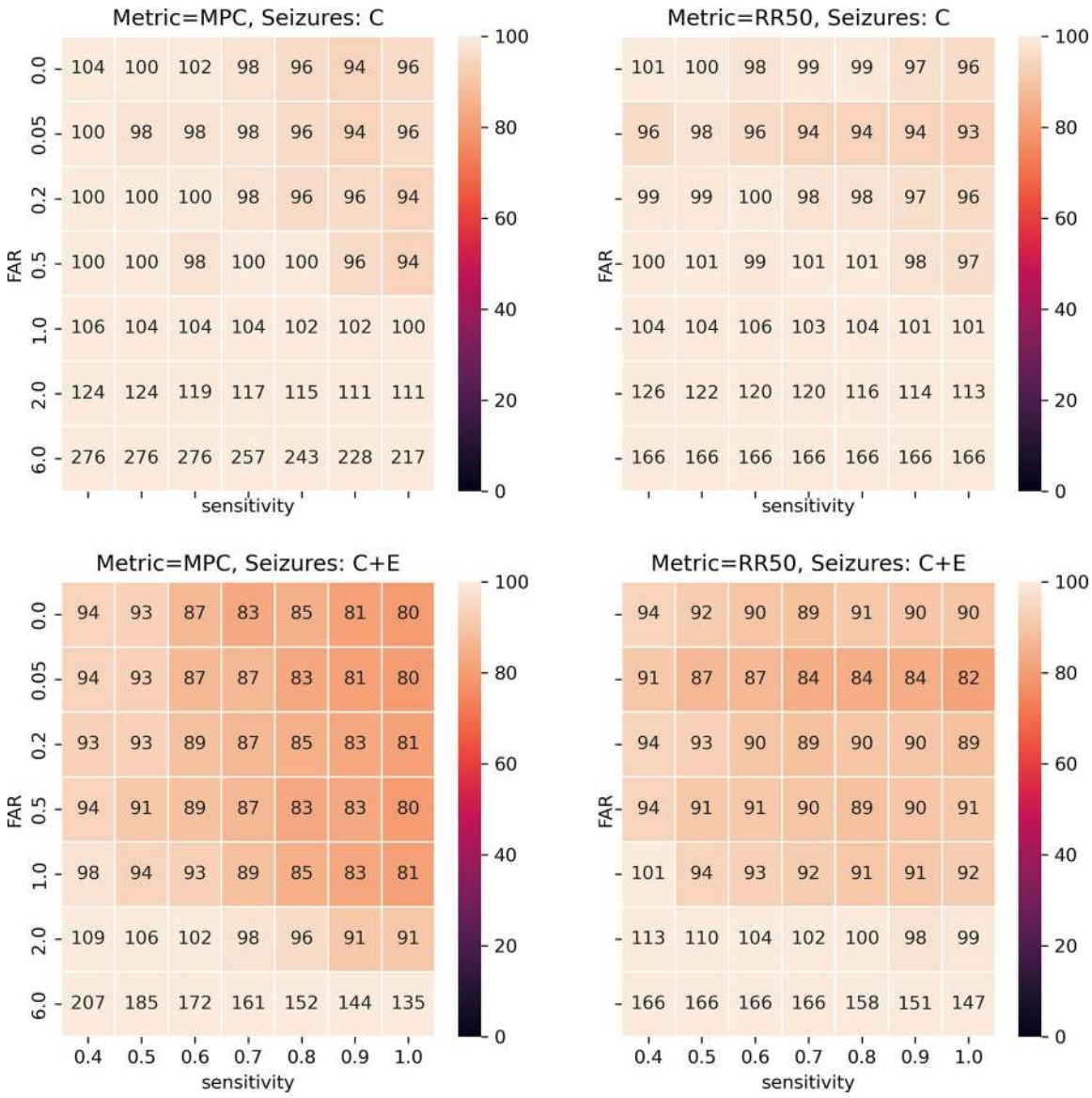

Figure 1:

RCT use case. This figure summarizes the RCT simulations by normalizing the data from Appendix C Table S2 and S3 to the clinical observed self-reported case (i.e., sensitivity=0.5, FAR=0) as 100% for each power metric modality separately. Thus, values below 100% are more efficient (in terms of needed number of patients to obtain 90% power) than clinically observed self-report, and values above 100% are less efficient. Seizures: C refers to wearables that can detect clinical seizures, Seizures: C + E refers to implanted wearables that can detect both clinical and electrographic seizures. MPC = median percentage change, RR50=50% responder rate. Seen here, devices for C+E seizures can be more efficient with a wider range of implant accuracies, but the devices for C seizures can also be more efficient with a range of device accuracies as well. The efficiency values between MPC and RR50 do not match perfectly, as RR50 is less statistically efficient for RCTs than MP. Note, the color bar upper limit is 100. Values higher than 100 are considered “undesirable” as the wearable

produced efficiency that is lower than self-report. The values are shown but all values above 100 are displayed with pale color.

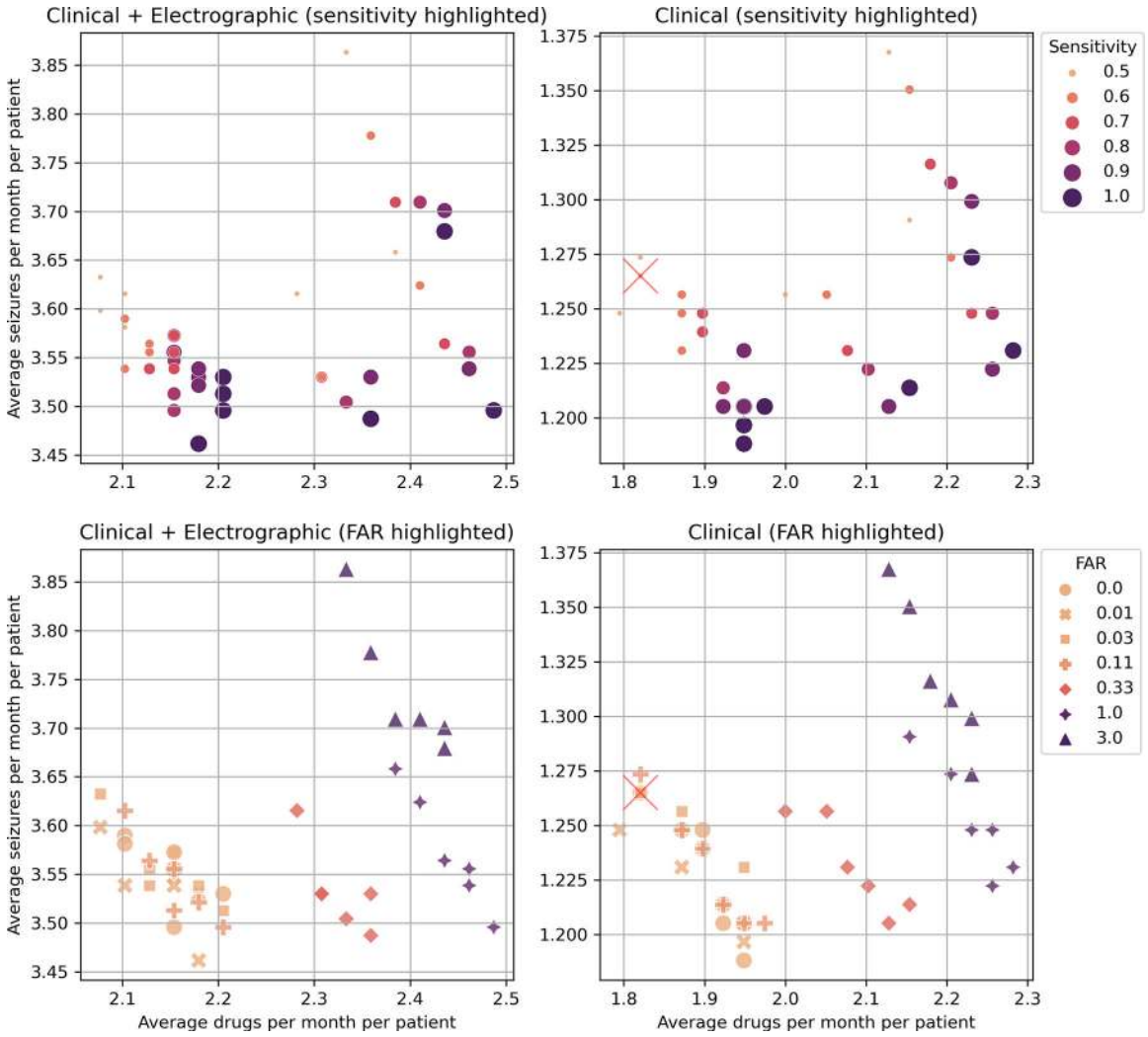

FIGURE 2: Medication adjustment in clinic use case.

Each point on these plots represents a summary of simulations for a specific type of wearable. Left column: implanted device case for clinical and electrographic seizures, right column: wearable device case for clinical seizures. Average drugs given per month per patient are plotted against average number of true seizures per month per patient. The upper and lower rows represent the same underlying simulated data but with different aspects highlighted. Upper row: Sensitivity is highlighted and is represented by radius of circle. Lower row: color and shape represent false alarm rate (FAR) which is highlighted. The lower right datapoints have higher sensitivity and upper left points have lower sensitivity. Similarly, lower FAR values are associated with farther lower left values, and higher FAR values are associated with farther upper right values. A red X is marked on the upper right and lower right plots for self-reported seizures (FAR=0, sensitivity=0.5).

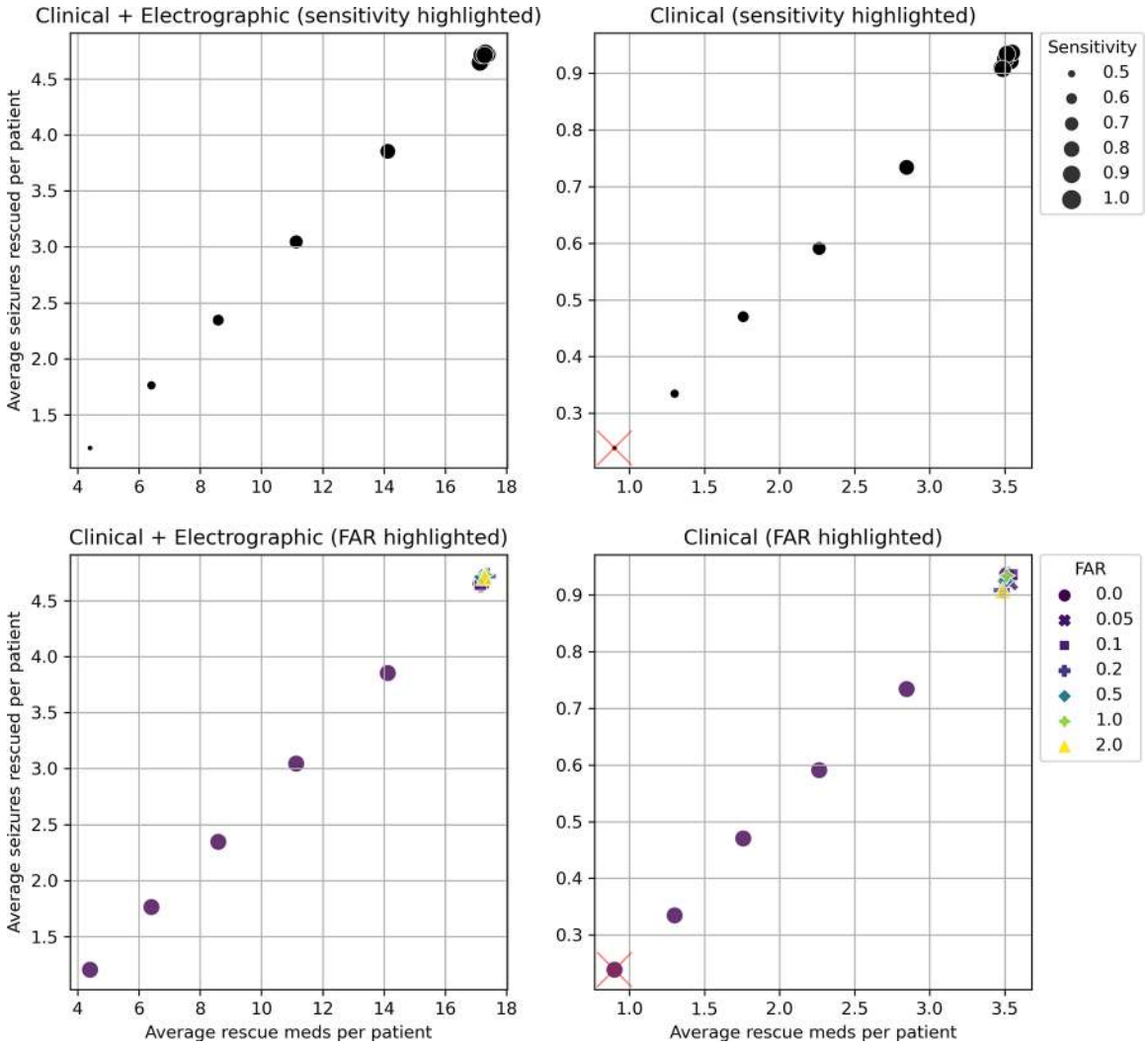

Figure 3: Cluster treatment use case.

Left column: clinical+electrographic seizures. Right column: clinical seizures. The same data is presented in the upper and lower rows with different aspects of the data highlighted. Upper row: sensitivity is highlighted via marker size. Lower row, false alarm rate (FAR) is highlighted via marker color and shape. The red X on the right upper and right lower plots indicates the self-reported seizure case (FAR=0, sensitivity=0.5). The x axis in all 4 subplots refers to the average (across 10,000 patients) number of rescue medications used per patient.

Table 1:

Injury prevention use case.

The injury counts from 10,000 simulated patients (10 years per patient). The detected injury rate closely matches the sensitivity of the device. The mean injury rate per patient is roughly 10 times smaller than the injury per 100-patient years (as expected). Seizure types are C (clinical) or C+E (clinical+electrographic). In all cases the false alarm rate equals 0 alarms/day. Injuries detected (%) = percentage of the total injuries detected by the wearable on average across patients. C – Clinical seizures only; C+E – clinical and electrographic seizures (implanted EEG device).

Seizure type	Simulated Injury rate per 100 patient-yrs	Sensitivity (%)	Injuries detected (%)	Injury-free (out of 10,000)
C	4.8	50	50	6297
C	4.8	60	60	6230
C	4.8	70	70	6231
C	4.8	80	80	6286
C	4.8	90	89	6222
C	4.8	100	100	6227
C	296	50	50	23
C	296	60	59	19
C	296	70	70	16
C	296	80	80	24
C	296	90	89	22
C	296	100	100	20
C+E	4.8	50	50	6196
C+E	4.8	60	59	6197
C+E	4.8	70	70	6278
C+E	4.8	80	80	6307
C+E	4.8	90	89	6255
C+E	4.8	100	100	6251
C+E	296	50	50	17
C+E	296	60	59	22
C+E	296	70	70	25
C+E	296	80	80	22
C+E	296	90	90	20
C+E	296	100	100	22

TABLE 2:

SUDEP prevention case.

SUDEP and aborted SUDEP events for the different wearable cases. Seizures=C for wearable capable of detecting clinical seizures, Seizures=C+E for implanted device capable of detecting clinical and electrographic seizures. SUDEP = number of simulated patients who "died" from an undetected SUDEP (out of 10,000). Aborted SUDEP = number of simulated patients who could have died from SUDEP but were prevented thanks to the wearable detecting and enabling the patient to get help in time. Prevented = the percentage of potential SUDEP events that were prevented by the wearable seizure detection. Sensitivity and the % prevented are overall closely aligned.

Seizures	Sensitivity	Prevented	SUDEP	Aborted SUDEP
C	50%	51%	900	955
C	60%	58%	759	1057
C	70%	69%	604	1326
C	80%	82%	346	1530
C	90%	91%	175	1770
C	100%	100%	0	1925
C+E	50%	50%	753	751
C+E	60%	62%	587	944
C+E	70%	71%	431	1051
C+E	80%	80%	294	1149
C+E	90%	91%	140	1334
C+E	100%	100%	0	1520