Epilepsia Open® The Open Access Journal of the International League Against Epilepsy

Open Access

FULL-LENGTH ORIGINAL RESEARCH

Characteristics of large patient-reported outcomes: Where can one million seizures get us?

*Victor Ferastraoaru, †Daniel M. Goldenholz, ‡Sharon Chiang, §Robert Moss, ¶William H. Theodore, and *Sheryl R. Haut

Epilepsia Open, 3(3):364–373, 2018
doi: 10.1002/epi4.12237

Dr Victor Ferastraoaru is Assistant Professor of Neurology at Albert Einstein College of Medicine and Montefiore Medical Center.

Summary

Objective: **To analyze data from Seizure Tracker, a large electronic seizure diary, including comparison of seizure characteristics among different etiologies, temporal patterns in seizure fluctuations, and specific triggers.**

Methods: **Zero-inflated negative binomial mixed-effects models were used to evaluate temporal patterns of seizure events (during the day or week), as well as group differences in monthly seizure frequency between children and adults and between etiologies. The association of long seizures with seizure triggers was evaluated using a mixed-effects logistic model with subject as the random effect. Incidence rate ratios (IRRs) and odds ratios were reported for analyses involving zero-inflated negative binomial and logistic mixed-effects models, respectively.**

Results: **A total of 1,037,909 seizures were logged by 10,186 subjects (56.7% children) from December 2007 to January 2016. Children had more frequent seizures than adults did (median monthly seizure frequency 3.5 vs. 2.7, IRR 1.26; $p < 0.001$). Seizures demonstrated a circadian pattern (higher frequency between 07:00 a.m. and 10:00 a.m. and lower overnight), and seizures were reported differentially across the week (seizure rates higher Monday through Friday than Saturday or Sunday). Longer seizures (>5 or >30 min) had a higher proportion of the following triggers when compared with shorter seizures: "Overtired or irregular sleep," "Bright or flashing lights," and "Emotional stress" ($p < 0.004$).**

Significance: **This study explored a large cohort of patients with self-reported seizures; strengths and limitations of large seizure diary databases are discussed. The findings in this study are consistent with those of prior work in smaller validated cohorts, suggesting that patient-recorded databases are a valuable resource for epilepsy research, capable of both replication of results and generation of novel hypotheses.**

KEY WORDS: **Big Data, Seizure, Electronic diary, Seizure trigger, Epilepsy fluctuation.**

Accepted May 12, 2018.

*Department of Neurology, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York, U.S.A.; †Division of Epilepsy, Beth Israel Deaconess Medical Center, Boston, Massachusetts, U.S.A.; ‡Department of Neurology, University of California San Francisco, San Francisco, California; §Department of Statistics, Rice University, Houston, Texas, U.S.A.; §SeizureTracker LLC, Alexandria, Virginia, U.S.A.; and ¶National Institutes of Health, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, U.S.A.

Address correspondence to Victor Ferastraoaru, Department of Neurology, Albert Einstein College of Medicine and Montefiore Medical Center, 111 E 210th Street, Bronx, NY 10467, U.S.A. E-mail: vferastr@montefiore.org

Seizure Tracker (www.seizuretracker.com) is a web-based and mobile app providing persons with epilepsy and their families an electronic diary for recording seizures, antiseizure drugs, and other important clinical data. The software includes >20,000 registered users and >1 million recorded seizures. Investigators in the International Seizure Diary Consortium[1] use exploratory techniques for Big Data[2] to investigate questions that cannot be approached with conventional designs. Data mining (data-driven hypothesis generation) can reveal insights about diseases that might be otherwise undetectable, as some patterns are revealed only when viewed in aggregate. Data-mining techniques have

Key Points

- Data from large electronic seizure diary databases have limitations but yield important insights into seizure patterns
- The temporal distribution of seizures demonstrated a circadian pattern and a larger number of seizures reported Monday through Friday
- Children had more frequent seizures than adults did, often even for the same epilepsy etiology
- Longer seizures (>5 or >30 min) were more likely to have the following triggers: "Overtired or irregular sleep," "Bright or flashing lights," or "Emotional stress"

been applied to large databases for several neurologic conditions, including epilepsy.[3,4] The Seizure Tracker database represents a rare opportunity to evaluate an extensive range of ages and etiologies, including uncommon forms of epilepsy.

A number of studies have been conducted using the SeizureTracker.com database previously[5–12]; however, none have explored the basic characteristics of the patient population using the database. Some commentators have previously considered the user base of this software to be all children, or extreme cases only. Given the variety of publications based on this database, it is important to provide a more in-depth characterization of the user base, the seizure types represented, and any baseline patterns noted.

In this study, we explored seizure characteristics for different etiologies among children and adults, fluctuation of seizures, and specific triggers. Our data illustrate the strengths and limitations of an online seizure diary. Studies analyzing Big Data could offer alternatives and supplements to standard clinical studies, with the potential to provide, in addition to their clear advantages of scale, longer-term and more widely population-based data.

Methods

Database structure and pre-processing

The Seizure Tracker database has been described.[5] Data were obtained in accordance with the National Institutes of Health (NIH) Human Research Protection Program (OHSR#12301). Briefly, data for subjects who agreed (opt-out alternative) were de-identified and unlinked. Subjects with no date of birth recorded and/or for whom only one event was reported (typically a test entry) were removed. Seizure times were treated as missing if the time of onset was not consistent with database dates, if they were duplicates, if a default time was recorded (01:00 a.m.), or if a nonpositive seizure duration was recorded. For analyses involving longitudinal measurements (seizure frequency, interseizure interval, circadian rhythm and weekly fluctuations, seizure clustering), diary spans of <30 days were excluded in order to increase reliability.

Data analysis

Statistical analysis was performed using R version 3.1.3 (Vienna, Austria). To account for repeated measures, generalized linear mixed-effects models with the subject as the random effect were used, with canonical link names as specified below. Seizure count data have been found to be empirically overdispersed relative to that expected under Poisson models; furthermore, seizure diary data are often zero-inflated, containing a larger number of zeros than under traditional Poisson or negative-binomial models. The zero-inflated negative binomial model provides a method for capturing these characteristics, by accounting for zero-inflation[7,8,13] as well as including an additional parameter to allow for overdispersion.[7,10,14] Group differences in monthly seizure frequency between children (<18) and adults (18 and older) were therefore compared using a zero-inflated negative binomial mixed-effects model with adulthood status during the month of the seizure as the fixed effect and subject as the random effect. Adulthood status during each month was assigned based on the patient's age at the start of the month; that is, for children who became adults partway through the month, this was attributed as a "childhood" seizure frequency. Etiologies with significantly higher seizure frequency were similarly identified through a zero-inflated negative binomial mixed-effects model, with etiology as the fixed effect and subject as the random effect.

To evaluate the temporal pattern of seizure events, a histogram of the frequency of seizure events from all patients against time was plotted for hours of the day. Midnight (12:00 a.m.) to 12:59:59 a.m. was defined as hour 0. Significant changes in seizure counts across time were identified through a zero-inflated negative binomial mixed-effects model, with subject as the random effect and hour-of-day as the fixed effect.[15] To estimate the amount of bias induced by treating default recorded times (01:00 a.m.) as missing, cubic spline interpolation was used to compute the estimated number of seizure events at 01:00 a.m. The hour with the lowest seizure rate was used for the reference hour, as suggested by Seneviratne et al. (2016). Temporal associations of seizure count with day-of-week were evaluated similarly, with the day with the lowest seizure rate, Sunday, used as the reference day.

Status epilepticus (long seizure) was defined using both the prior definition of 30-minute time limit as well as a more recent 5-minute time limit. The association of status epilepticus with seizure triggers was evaluated using a mixed-effects logistic model with subject as the random effect. Failure to report any trigger was assumed to be data Missing Completely at Random and were also excluded. Operational definitions of seizure clusters vary. For the purpose of this study we used the clustering definition of 3 or more

consecutive seizures in any given 24 hours[16] and analyzed clustering for patients who had <1,000 seizures (who had discrete seizures).

For all analyses, statistical significance was assigned at the $\alpha = 0.05$ level, with false discovery rate control through the Benjamini–Hochberg procedure. Due to the large sample size, effect sizes are reported for all statistically significant results through the incidence rate ratio (IRR) for analyses involving zero-inflated negative binomial mixed-effects models or the odds ratio (OR) for analyses involving logistic mixed-effects models.

Results

Demographics

Data included all entries between December 2007 and January 2016, comprising 22,806 subjects and 1,123,600 seizures. After pre-processing, 12,615 subjects were eliminated (of whom 8,203 had no seizures recorded), resulting in 10,186 subjects and a total of 1,037,909 seizures for analysis. Diary durations extended up to more than 8 years (median 82.7 days, interquartile range [IQR] 367.0 days). A total of 3,493 subjects had diary durations <30 days and were excluded from longitudinal analyses (seizure frequency, interseizure interval, circadian rhythm, and weekly fluctuations).

Fifty-two percent were female, with 45% male (gender not listed for 2.7%). This included 43.3% adults (18 and older) and 56.7% children (<18) at the time of initial diary entry, with an additional 1.7% of patients becoming adults by the time they had the most recent seizure recorded in the database. Epilepsy "etiology" was listed for 42.5% of subjects; some listed more than a single etiology. "Brain Trauma" was the most frequent etiology overall at 8.7%. The 6 most common etiologies reported for children were "Genetic Abnormalities" (324 children; this represents 3.18% of the entire population studied and 7.1% of all children), "Brain Malformations" (5.6% of all children), "Tuberous Sclerosis" (5.2% of all children), "Brain Trauma" (5.0% of all children), "Dravet Syndrome" (4.8% of all children), and "Lack of oxygen during birth" (3.5% of children). For adults, these were "Brain Trauma" (606 patients, this represents 5.95% of the entire population studied and 13.2% of all adults), "Infection" (6.3% of all adults), "Brain Tumors" (4.3% of all adults), "Genetic Abnormalities" (3.5% of all adults), "Brain Malformations" (3.8% of all adults), and "Stroke" (2.7% of all adults). Detailed demographic data are presented in Table 1.

Seizure type, duration, frequency, and interseizure intervals

Eighty-five percent of seizures were classified as one of the seizure types listed in Table 1, and 14.6% were reported as "Unknown" or "Other". The most frequent seizure types were the same for children and adults: Focal aware, Focal impaired awareness, and [Unknown] onset tonic–clonic. The seizure terminology used was adapted using the 2017 Operational classification of seizure types by the International League Against Epilepsy.[17,18] Detailed seizure type data, including distribution among children and adults and median seizure durations, are presented in Table 1. The distribution of each seizure type is illustrated in Figure 1 for each etiology.

Seizure duration ranged from 1 second to 1 day for the 751,625 seizures with recorded duration, with a median of 30 seconds (IQR, 95 s). More than 90% of the seizures lasted 5 minutes or less; 1.5% were reported to have duration >30 minutes.

Overall, children had a higher seizure frequency than adults (IRR 1.26, $p < 0.001$), with a mean (median) seizure frequency of 16.1 (3.5) seizures per month among children and 7.7 (2.7) seizures per month among adults, respectively (Fig. S1). In adults, increased seizure frequency was reported for patients with Lennox–Gastaut syndrome (IRR, 1.24; $p_{uncorrected} = 0.01$), although this difference did not remain statistically significant after multiple testing correction. In children, significantly higher seizure frequency was reported in patients with Aicardi syndrome (IRR, 3.98; $p < 0.001$). Children with Lennox–Gastaut syndrome (IRR, 1.16; $p_{uncorrected} = 0.037$) and Tuberous Sclerosis Complex (IRR, 1.14; $p_{uncorrected} = 0.040$) also reported significantly increased seizure frequencies, although associations with Lennox-Gastaut syndrome and tuberous sclerosis complex did not remain statistically significant after multiple testing correction.

For patients who had diary duration of 30 days or more, the interseizure intervals (ISIs) >0 were analyzed. The median ISI was 4 hours (IQR 23 h) and the mean ISI was 77.4 hours (STD 662.7 h). Of the total ISIs, 40.2% were 2 hours or less, 9.4% were between 2–4 hours, 8.6% were 4–8 hours, 18.9% were 8–24 hours, 21.4% were 1–30 days, and 1.6% were >30 days (Fig. S2). Of these patients, 62.5% had at least one ISI of 4 hours or less and 9.7% had a median ISI of 4 hours or less.

Circadian and weekly fluctuation of reported seizures

The temporal distribution of seizures demonstrated a circadian pattern. The estimated total number of seizure events at 01:00 a.m. based on cubic spline interpolation was 2,871. The lowest seizure rate occurred at hour 01:00 a.m. to02:00 a.m., both before and after cubic spline interpolation, with 13,063 observed seizure events when events recorded as 01:00 a.m. were treated as missing, and 15,934 seizure events when the interpolated 2,871 seizure events were included. The IRR, which estimates the association between each hour of the day and the mean number of seizures per hour per person, provides a measure of effect size and is shown in Table S1. Seizure rates were highest early in the day, peaking at hour 07:00 a.m., and lowest overnight with a decline after hour 06:00 p.m. (Fig. 2A,B; Table S1).

Table 1. Demographic data and seizure characteristics of patients in Seizure Tracker database

	Value (median) (IQR)		
Age in years[a]	15.8 (24.8)		
	Children	Adults	
Length of seizure diary in days[b,c]	80 (359)	82 (386)	
Number of total seizures recorded[d]	12 (52)	8 (28)	
Seizure frequency per month[e]	3.5 (10.1)	2.7 (5.7)	
Seizure duration in seconds	30 (81)	30 (112)	
Interseizure interval in hours[f]	3.0 (18.5)	9.5 (43.5)	
	Value (% of total population)		
Gender	Female (52.3), Male (45), Unknown (2.7)		
Age groups in years[a]	0–2 (5.1), 2–10 (29.4), 10–18 (20.5), 18–40 (29.7), 40–60 (13.7), 60–85 (1.6)		
Epilepsy etiology[a,g]	Children	Adults	
Aicardi syndrome[h]	0.34	0.08	
Angelman's syndrome	0.23	0.07	
Down's syndrome	0.25	0.17	
Dravet syndrome	2.68	0.23	
Lennox-Gastaut syndrome	1.63	0.72	
Neurofibromatosis	0.11	0.09	
Rett syndrome[h]	0.50	0.13	
Sturge-Weber syndrome	0.20	0.05	
Tuberous sclerosis	2.87	0.67	
Brain tumors	1.15	1.93	
Brain trauma	2.80	5.95	
Infection	1.84	2.84	
Stroke	1.51	1.22	
Lack of oxygen during birth	1.96	1.16	
Maternal drug or alcohol abuse	0.27	0.23	
Alcohol or drug abuse	0.07	0.45	
High fever	1.01	1.05	
Genetic abnormalities	3.18	1.59	
Brain malformations	3.05	1.73	
Other[j]	2.40	3.49	
	Value (% of all seizures)		**Duration in seconds (median) (IQR)**
Seizure type[j,k]	Children	Adults	
Focal aware (former "Simple partial")	11.2	6.5	20 (52)
Focal impaired awareness	9.9	5.5	50 (100)
[Unknown] onset tonic–clonic	10	4.5	60 (97)
[Focal/generalized] tonic	8	3.6	10 (29)
[Focal/generalized] myoclonic[l]	7	1	10 (117)
Absence and absence, atypical	5.5	1.9	15 (55)
[Focal/generalized] atonic	3.5	0.3	10 (34)

Continued

For example, the IRR of 5.36 at hour 07:00 a.m. indicates that the number of reported seizures was expected to be 5.36 times higher than the seizure frequency at hour 01:00 a.m. A mild increase in the number of reported events occurred between hour 12:00 p.m. and 18:00 p.m., an effect which was attenuated when hourly seizure clusters (defined as multiple, i.e., t2 or more, seizures occurring within 1 hour) were treated as single events (Fig. 2C,D). Seizure rates were higher Monday through Friday than Saturday or Sunday (Fig. 3; Table S2). This weekly pattern was present both before and after daily seizure clusters (defined as 3 or more seizures occurring within the same day)[19–23] were treated as single events (Fig. 3C,D).

Seizure triggers

Triggers were reported for 32.2% (334,601) of reported seizures. The most frequent triggers reported were "Overtired or irregular sleep" (14.2%), followed by "Other" (12.5%), "Changes in medication" (6.3%), "Hormonal fluctuations "(5.5%) and "Emotional stress" (5.0%).

The relative proportion of triggers for seizures that lasted 5 minute or shorter in duration (233,611) was compared with seizures longer than 5 minutes (28,134). Seizures were more likely to last more than 5 minutes if triggered by being "Overtired or irregular sleep" (OR 1.30, $p < 0.001$), "Diet" (OR 1.17; $p = 0.002$), "Bright or flashing lights" (OR 1.21; $p = 0.002$), "Emotional stress" (OR 1.25, $p < 0.001$), or

Table 1. Continued.

	Value (median) (IQR)		
[Focal/generalized/unknown] onset epileptic spasms[l]	2.3	0.06[m]	240 (391)
Focal to bilateral (tonic–clonic)	1.5	0.7	60 (100)
Focal aware (former "Aura only)	0.4	1.2	10 (40)
[Focal/generalized] clonic	0.4	0.3	25 (110)
Focal [aware or impaired awareness] emotional	0.07	0.03	50 (106)

[a]Age at most recent seizure diary recording.

[b]Additional data: 2.3% of patients reported catamenial data, 50.5% reported at least twice, data regarding medication use (dose, duration), 1.2% reported medication blood levels, 23.3% reported their weight.

[c]Range <1 day to 8.1 years (median of 82.7 days, IQR 367 days); 64.8% of seizure diaries were >1 month and 26.1% >1 year.

[d]Range 2–33,033 seizures (median 10 seizures, IQR 41 seizures).

[e]Mean (standard deviation) 16.1 (46.5) in children, 7.7 (20.0) in adults.

[f]Mean (standard deviation) 62.0 (577.9) in children, 111.0 (808.3) in adults.

[g]Results reported here as % of total population. A total of 42.5% of all subjects listed an "etiology" for epilepsy; some listed more than a single etiology. Most frequent etiologies overall: "Brain Trauma" (8.75%), "Tuberous Sclerosis" (3.54%).

[h]Almost all patients in this category were female.

[i]Etiologies not presented here were vague or had very few patients. The majority were listed as "Other." The additional etiologies listed were: "hypothalamic hamartoma" (with a total 5 patients), "Phelan-McDermid syndrome"(4 patients), "Alzheimer's" (9 patients), "heart attack" (19 patients), "brain surgery" (274 patients), "metabolic disorder" (80 patients), "electrolyte disturbances" (49 patients), "brain injury during fetal development" (221 patients), "lead exposure" (18 patients), "carbon monoxide exposure" (10 patients).

[j]Seizure types were adapted according to 2017 Operational classification of seizure types by the International League Against Epilepsy: "Simple partial" to Focal aware, "Complex partial" to Focal impaired awareness, "Tonic Clonic" to [Unknown] onset tonic–clonic, "Tonic" to [Focal/generalized] tonic, "Myoclonic & Myoclonic cluster" to [Focal/generalized] myoclonic, "Absence and Atypical Absence" to Absence and Absence, atypical, "Atonic" to [Focal/generalized] atonic, "Infantile spasms (cluster)" to [Focal/generalized/unknown] onset epileptic spasms, "Secondarily Generalized" to Focal to bilateral (tonic–clonic), "Aura Only" to Focal aware, "Clonic" to [Focal/generalized] clonic, "Gelastic" to Focal [aware or impaired awareness] emotional.

[k]14.6% of total seizures were reported as "Unknown" (e.g., Unclassified) or "Other." In addition, other seizure characteristics were reported. A total of 9.7% (100,650) seizures had auras. Clinical manifestations were documented for 78.3% of total seizures (20% reported change in awareness, 17.9% reported loss of ability to communicate, and 69.1% reported motor manifestations such as muscle stiffness, muscle twitching). Postictal phase was reported in 32.2%, comprising either "Unable to communicate" (8.2%), "Muscle weakness," (7.9%) or "Sleepy" (23.3%).

[l]These were listed either as clusters or as duplicates.

[m][Focal/generalized/unknown] onset epileptic spasms in adults analyzed as error.

having "Fever or overheated" (OR 1.35; $p < 0.001$). Seizures lasting 30 minutes or shorter in duration (256,323) were also compared with seizures lasting longer than 30 minutes (5,422). Seizures were more likely to last more than 30 minutes if triggered by "Change in medications" (OR 1.19; $p = 0.003$), "Overtired or irregular sleep" (OR 1.17, $p = 0.003$), "Bright or flashing lights" (OR 1.42, $p = 0.003$), or "Emotional stress" (OR 1.249, $p = 0.0009$).

Seizure clusters

For patients who had diary duration of 30 days or more, the seizure-clustering patterns were analyzed: 54.6% of seizures were part of a cluster and 55.7% of the patients had at least one seizure cluster (44.3% of patients did not have any seizure cluster). Seizure clustering pattern for each etiology is displayed in Figure 4.

Discussion

Online seizure diary as a tool

In this study we examined seizure patterns and seizure characteristics from a large online database. This study provides insight into how exploratory analysis of self-reported Big Data can help elucidate phenomena such as seizure fluctuations and better characterize epilepsy variables. Although a number of epilepsy studies have reported on the temporal distribution of seizures from prospective diaries,[24,25] a Big Data approach provides access to larger sample sizes and prolonged reporting periods.

The primary observational findings of this study are listed below:

Circadian variations emerged clearly from the data—more seizures were reported during morning hours as compared to overnight (peak of reported seizures at 07:00 a.m.), which replicate several previous epilepsy studies.[26] Using intracranial electroencephalography (EEG) data captured with a Responsive Neurostimulator System (RNS), Duchrow reports a bimodal daily variation of seizures distribution, with relative maxima at 06:00 and 15:00 hours.[27] In a different study of ambulatory intracranial monitoring (RNS), Spencer analyzes circadian and ultradian patterns of epileptiform discharges and how they differ by seizure-onset location and demonstrates that cyclic occurrence of epileptiform discharges and seizures are influenced by the cumulative effects of various circadian rhythms that vary in influence by the pathophysiology of the underlying epilepsy syndrome.[28] In our study population, we found a novel pattern, with fewer seizures reported at the end of the week (Saturday and Sunday). This interesting finding may reflect the proposed association between stress and seizure occurrence.[29]

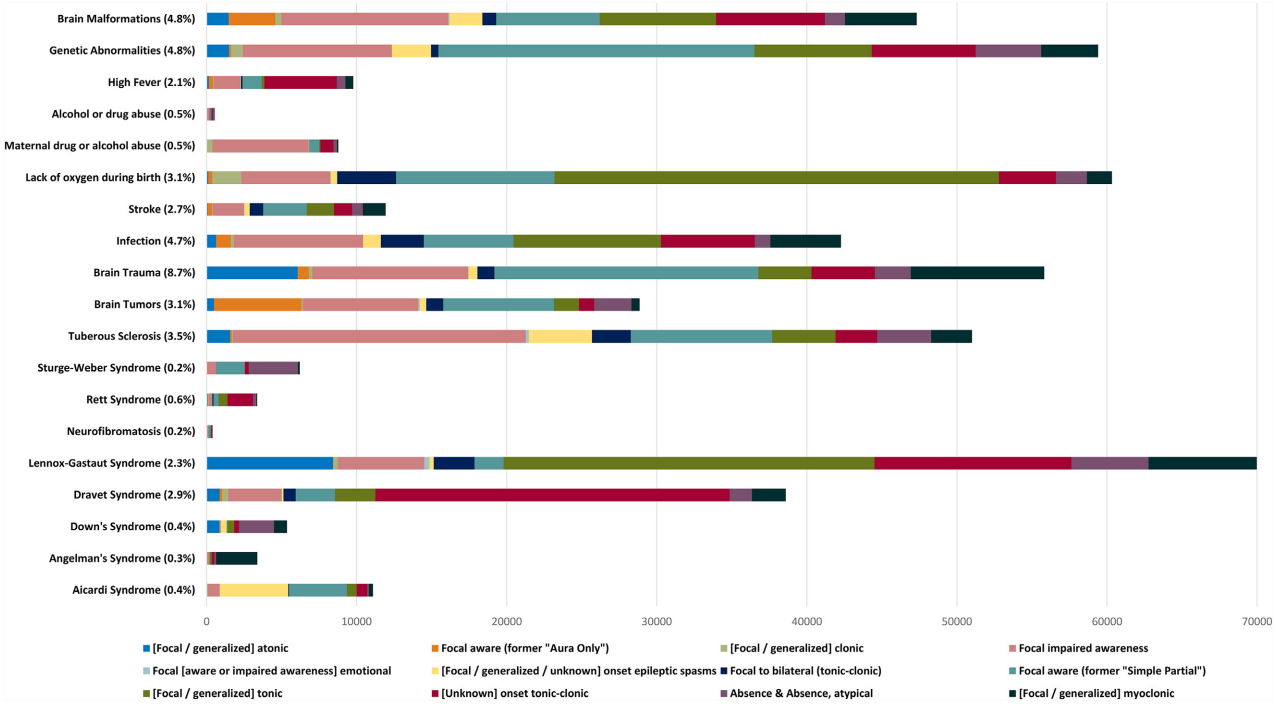

Figure 1.
Distribution of seizure types across reported epilepsy etiologies.[1,2]
[1]On the x axis, the numbers represent the overall number of seizures reported; on the y axis each etiology is listed (proportions in parenthesis represent % of total population). [2]Note for example, that for patients with "Brain Trauma" or "Brain Tumor," the most frequent seizure types were Focal impaired awareness and Focal aware. This was the same for patients with "Tuberous Sclerosis," who additionally had frequent [Focal/generalized/unknown] infantile spasms. For the 238 patients with "Lennox–Gastaut syndrome," the most frequent seizure type reported was [Focal/generalized] tonic (33.3%), with several other seizure types being common: [Unknown] onset tonic clonic (17.7%), [Focal/generalized] atonic (11.3%), and [Focal/generalized] myoclonic (9.7%). The most frequent seizure type for patient with "Aicardi syndrome" were [Focal/generalized/unknown] onset epileptic spasms (37% of all their seizures) and for patients with "Angelman syndrome," these were the [Focal/generalized] myoclonic seizures (74%).

2 Children had more frequent seizures than adults did, often even for the same epilepsy etiology. This suggests that the natural history of epilepsy related to certain etiologies may change with aging.

3 The most frequently reported triggers were similar to previous cross-sectional[30] and prospective diary studies,[31] including sleep-related issues, medication changes, and emotional stress. Of interest, triggers varied by seizure duration, such as "Overtired or irregular sleep," "Bright or flashing lights," or "Emotional stress" were associated with long seizures/status epilepticus. The role of seizure triggers or precipitants remains an increasing area of interest in predicting seizures and understanding the consequences of prolonged seizures.[32] These are also important in the emerging field of epilepsy self-management.[33]

4 Many inter-seizure intervals (or ISIs) were brief, likely related to the large number of epilepsy etiologies with tendency to cluster as shown in Figure 4. Previous studies have reported variable prevalence of seizure clustering 7–76%, often with higher prevalence of seizure clusters in relation to inpatient monitoring data, when medications are being decreased or stopped to capture seizures.[34]

Limitations of large patient-reported seizure diary databases

The use of an electronic patient-reported database presents challenges and limitations. The most important one is the reliability of the data. Epilepsy diaries are not electrographically verified. Recorded seizures may be nonepileptic, misclassified, or over- or underreported. In addition, perceived event durations may be inaccurate.[25] This challenge is currently present in nearly all studies that rely on prospective epilepsy diaries, including formal clinical trials.[35,36] We observed a small percentage (0.06% overall) of Focal/generalized/unknown onset epileptic spasms reported as coming from adults. Although this certainly represents an error, it is difficult to decide whether the error was attributable to an incorrect input regarding the date of birth,

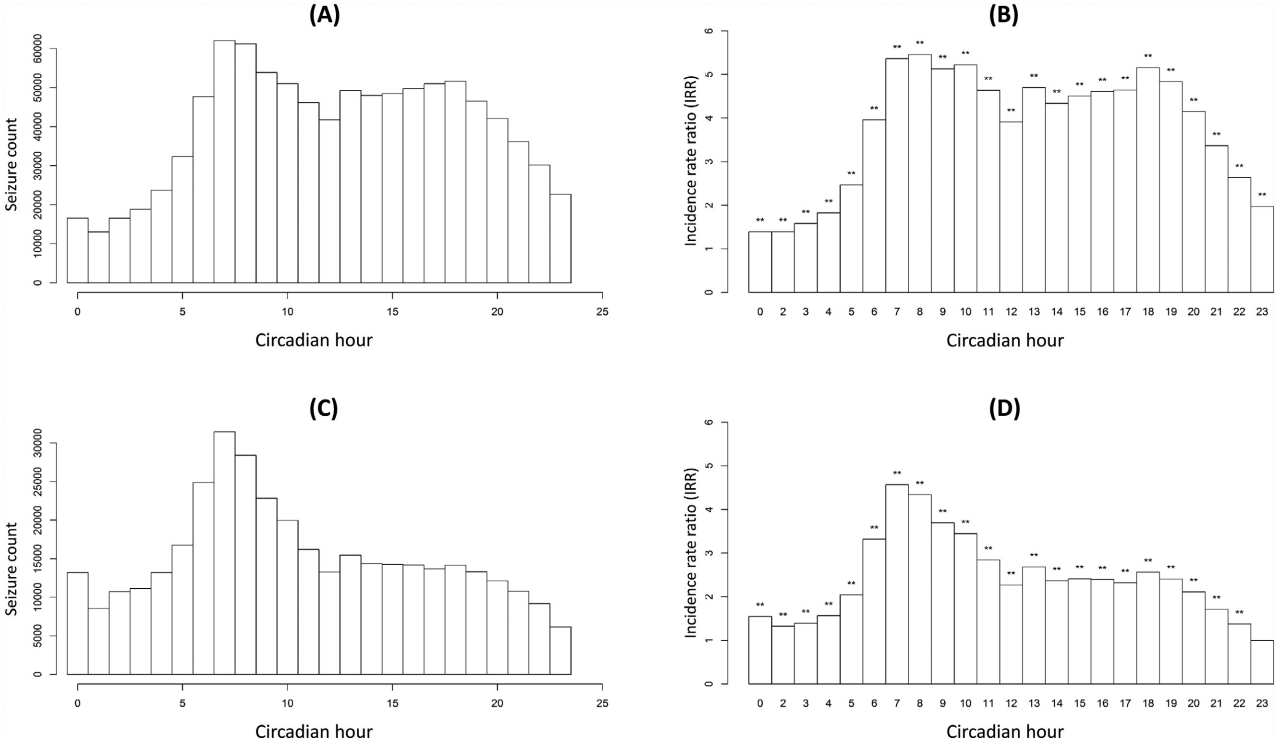

Figure 2.
(**A**) Histogram of the total number of seizures from all patients against time of day. (**B**) Comparison of seizure events for circadian hours with the reference hour (hour 01:00 a.m.) in mixed-effects negative binomial model of 9,849 patients. Incidence rate ratios (or IRRs) are calculated as the exponentiated coefficients. For example, an IRR of 5.457 during hour 08:00 a.m. means that the number of reported seizures is 5.457 times higher than at hour 01:00 a.m. An IRR at hour 01:00 a.m. is not provided as this is the reference hour. Midnight (12:00 a.m.) to 12:59:59 a.m. is defined as hour 0. (**C**) Histogram of the total number of seizures from all patients against time of day, with hourly seizure clusters (defined as 2 or more seizures occurring within the same hour) treated as single events. (**D**) Incidence rate ratios for seizure events across circadian hours, with hourly seizure clusters (defined as in C) treated as single events. IRR, incidence rate ratio; SE, standard error. $^{**}p < 0.001$

misclassification of seizure type by patients/families, or simply a recording mistake when pressing the app button for making a choice. Other sources of error from patient-reported data may stem from lack of physician verification, including seizure miscounting, difficulty of caregivers distinguishing between generalized and focal to bilateral seizure evolution, or inclusion or psychogenic nonepileptic seizures in the database. Even data self-reported by well-trained observers must be considered imperfect in accuracy, but the advantage of large databases is that these errors are generally likely to have a relatively modest impact.

Another example of the limitations of patient-recorded databases involves the preprocessing of seizure events with a default recorded time of 01:00 a.m. illustrated above. By default, Seizure Tracker records a seizure event time of 01:00 a.m., and it is not possible to distinguish missing event times from true occurrences at 01:00 a.m. Therefore, we treated 01:00 event times as missing and used cubic spline interpolation to estimate the number of events occurring at 01:00 a.m. In our case, the total number of seizure events occurring at 01:00 a.m. based on cubic spline interpolation was 2,871. Although in this particular dataset, the difference of 14.3% (corresponding to the difference between 13,063 and 15,934 seizures) did not affect the use of the 01:00 a.m.–02:00 a.m. hour as the reference hour in circadian analysis, slight upward bias is likely in the incidence rate ratios shown in Figure 2. Although statistical significance of these results is unlikely to be affected due to the large sample size of the dataset, the clinical significance of circadian fluctuations from midnight to 01:00 a.m. and from 02:00 a.m. should be considered. It is possible that seizure underreporting is not a uniformly random process, but that a systematic bias

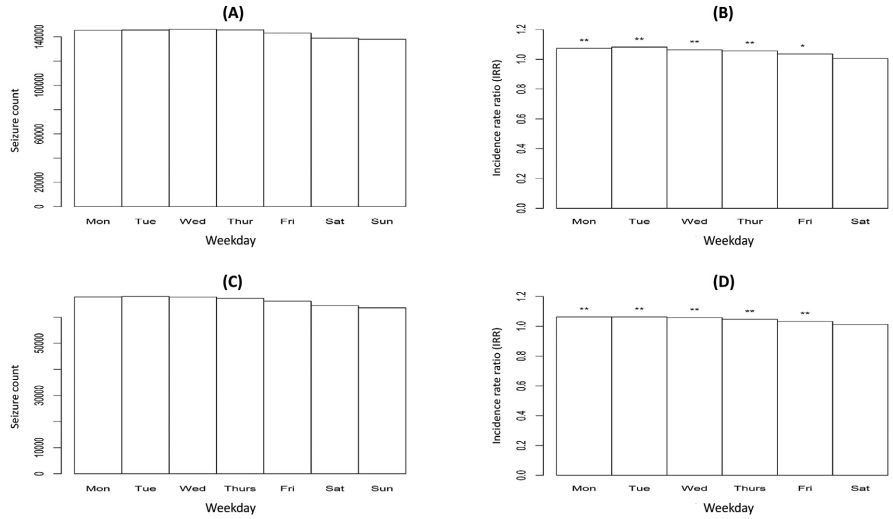

Figure 3.

(**A**) Histogram of the total number of seizures from all patients against day of week. (**B**) Comparison of seizure events for days of the week with the reference day (Sunday) in mixed-effects zero-inflated negative binomial model of 9,849 patients. Incidence rate ratios (IRRs) are calculated as the exponentiated coefficients. For example, an IRR of 1.074 on Monday means that the number of reported seizures is expected to be 7.4% higher than on Sunday. An IRR on Sunday is not provided as this is the reference day. (**C**) Histogram of the total number of seizures from all patients against day of week, with seizure clusters (defined as three or more seizures occurring within the same day) treated as single events. (**D**) Incidence rate ratios for seizure events against day of week, with seizure clusters (defined as in C) treated as single events. IRR, incidence rate ratio; SE, standard error. $^{**}p < 0.001$, $^{*}p < 0.01$

exists. In that circumstance our results would reflect that implicit bias. We have no reason to consider the systematic bias in these data to be fundamentally different from the underreporting of seizures in clinic patients or in randomized controlled trials.

Another key issue in the interpretation of results produced from analyses of Big Data involves the difference between statistical and clinical significance. Due to the asymptotic behavior of p-values under the null hypothesis, the large sample sizes employed in Big Data result in increased power, so that even miniscule differences become clinically significant. In such cases, it is important to consider not only statistical significance, but also clinical significance. In these cases, the reporting of effect sizes through measures such as Cohen's D, IRRs, or ORs may be more useful for interpretation. It is important to note that the patterns reported based on fixed effects represent population-based effects, which masks patient-specific variation. There may be individual clusters of patients who exhibit patterns that differ from the overall population-based effect. As an example, the estimated across-subject standard deviation in estimating circadian and weekly seizure fluctuation patterns was quite large relative to the magnitude of the largest fixed effect, indicating that the subjects varied widely relative to the magnitude of the circadian and weekly fixed effects (Tables S1 and S2).

Seizure Tracker also appears to be highly utilized by patients with severe epilepsy and frequent seizures, which may limit generalizability. For example, patients were noted to report high monthly seizure frequencies and short interseizure intervals. In addition, a large proportion of patients reported Aicardi syndrome (approximately 1% of all patients who reported an etiology in the database). Although this is a limitation, it is also a strength in that previously underrepresented populations in epilepsy research are now more available for study through these databases.

Despite their limitations, there is clearly much to learn from exploration of patient-reported databases in epilepsy. They provide access to much larger datasets, across longer time frames and wider populations than are available through standard clinical studies. Some of the findings here mirror those of smaller studies with high reliability, underscoring that patient-recorded databases still can make valid population level observations. Seizure Tracker is an important research and clinical tool for neurologists and

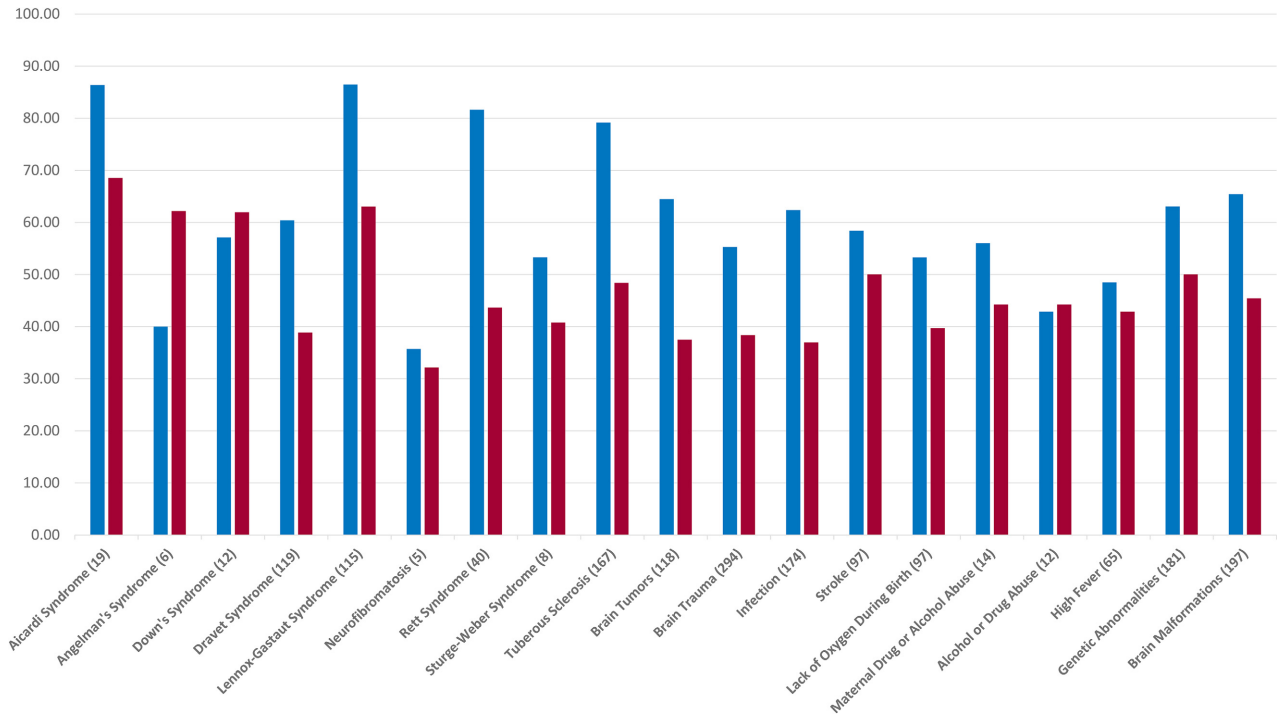

Figure 4.
Seizure clustering for each etiology (in parenthesis, number of patients who had at least one seizure cluster)*·**. *Blue: Proportion of patients with at least one seizure cluster. **Red: For patients who had at least one seizure cluster, the proportion of their seizures that were part of clusters varied between 1% and 100%. Displayed here is the median proportion of clustered seizures, for patients who had seizure clusters.

neuroscientists and provides a framework that can be employed to follow other neurologic conditions in the outpatient setting. The novel results presented here begin to expand our understanding of variation of seizures rates (e.g., during the day and week, in relation to certain triggers), which could ultimately lead to better planning of clinical trials and personalized patient treatments.[9]

Acknowledgments

The initial pre-processing of data performed at NIH was supported by the National Institutes of Neurological Disease and Stroke Intramural Research Program. Use of data was facilitated by the International Seizure Diary Consortium (https://sites.google.com/site/isdchome/).

Disclosure of Conflict of Interest

S.R. Haut served as consultant for Otsuka, Engage, and Eisai; R. Moss is the cofounder/owner of Seizure Tracker and received personal fees from Cyberonics, UCB, and Courtagen, and grants from Tuberous Sclerosis Alliance. V. Ferastraoaru, D.M. Goldenholz, S. Chiang, and W.H. Theodore have no conflicts of interest to disclose. 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. Preliminary results were presented at the 70th Annual American Epilepsy Society Meeting.

References

1. The International Seizure Diary Consortium (ISDC). Available at: https://sites.google.com/site/isdchome/. Accessed 2018.
2. Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: epidemiology in the era of big data. *Epidemiology* 2015;26:390–394.
3. Wagenaar JB, Worrell GA, Zachary I, et al. Collaborating and sharing data in epilepsy research. *J Clin Neurophysiol* 2015;32:235–239.
4. Devinsky O, Dilley C, Ozery-Flato M, et al. Changing the approach to treatment choice in epilepsy using big data. *Epilepsy Behav* 2016;56:32–37.
5. Goldenholz DM, Moss R, Scott J, et al. Confusing placebo effect with natural history in epilepsy – a big data approach. *Ann Neurol* 2015;78:329–336.
6. Goldenholz DM, Tharayil JJ, Kuzniecky R, et al. Simulating clinical trials with and without intracranial EEG data. *Epilepsia Open* 2017;2:156–161.
7. Tharayil JJ, Chiang S, Moss R, et al. A big data approach to the development of mixed-effects models for seizure count data. *Epilepsia* 2017;58:835–844.
8. Chiang S, Vannucci M, Goldenholz D, et al. Epilepsy as a dynamic disease A model for differentiating seizure risk from natural variability. *Epilepsia Open* (in press 2018).
9. Goldenholz DM, Goldenholz SR, Moss R, et al. Does accounting for seizure frequency variability increase clinical trial power? *Epilepsy Res* 2017;137:145–151.
10. Goldenholz DM, Goldenholz SR, Moss R, et al. Is seizure frequency variance a predictable quantity? *Ann Clin Transl Neurol* 2018;5:201–207.
11. Goldenholz DM, Strashny A, Cook M, et al. A multi-dataset time-reversal approach to clinical trial placebo response and the relationship to natural variability in epilepsy. *Seizure* 2017;53:31–36.

12. Goldenholz DM, Tharayil J, Moss R, et al. Monte Carlo simulations of randomized clinical trials in epilepsy. *Ann Clin Transl Neurol* 2017;4:544–552.
13. Hopkins A, Davies P, Dobson C. Mathematical models of patterns of seizures their use in the evaluation of drugs. *Arch Neurol* 1985;42:463–467.
14. Balish M, Albert PS, Theodore WH. Seizure frequency in intractable partial epilepsy: a statistical analysis. *Epilepsia* 1991;32:642–649.
15. Seneviratne U, Boston RC, Cook M, et al. Temporal patterns of epileptiform discharges in genetic generalized epilepsies. *Epilepsy Behav* 2016;64:18–25.
16. Haut SR. Seizure clusters: characteristics and treatment. *Curr Opin Neurol* 2015;28:143–150.
17. Fisher RS, Cross JH, French JA, et al. Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 2017;58:522–530.
18. Fisher RS, Cross JH, D'Souza C, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. *Epilepsia* 2017;58:531–542.
19. Haut SR, Swick C, Freeman K, et al. Seizure clustering during epilepsy monitoring. *Epilepsia* 2002;43:711–715.
20. Rose AB, McCabe PH, Gilliam FG, et al. Occurrence of seizure clusters and status epilepticus during inpatient video-EEG monitoring. *Neurology* 2003;60:975–978.
21. Haut S, Lipton R, LeValley A, et al. Identifying seizure clusters in patients with epilepsy. *Neurology* 2005;65:1313–1315.
22. Noe KH, Drazkowski JF. Safety of long-term video-electroencephalographic monitoring for evaluation of epilepsy. *Mayo Clin Proc* 2009;84:495–500.
23. Di Gennaro G, Picardi A, Sparano A, et al. Seizure clusters and adverse events during pre-surgical video-EEG monitoring with a slow anti-epileptic drug (AED) taper. *Clin Neurophysiol* 2012;123:486–488.
24. Fisher RS, Bartfeld E, Cramer JA. Use of an online epilepsy diary to characterize repetitive seizures. *Epilepsy Behav* 2015;47:66–71.
25. Fisher RS, Blum DE, DiVentura B, et al. Seizure diaries for clinical research and practice: limitations and future prospects. *Epilepsy Behav* 2012;24:304–310.
26. van Campen JS, Valentijn FA, Jansen FE, et al. Seizure occurrence and the circadian rhythm of cortisol: a systematic review. *Epilepsy Behav* 2015;47:132–137.
27. Duckrow RB, Tcheng TK. Daily variation in an intracranial EEG feature in humans detected by a responsive neurostimulator system. *Epilepsia* 2007;48:1614–1620.
28. Spencer DC, Sun FT, Brown SN, et al. Circadian and ultradian patterns of epileptiform discharges differ by seizure-onset location during long-term ambulatory intracranial monitoring. *Epilepsia* 2016;57:1495–1502.
29. Novakova B, Harris PR, Ponnusamy A, et al. The role of stress as a trigger for epileptic seizures: a narrative review of evidence from human and animal studies. *Epilepsia* 2013;54:1866–1876.
30. van Campen JS, Jansen FE, Pet MA, et al. Relation between stress-precipitated seizures and the stress response in childhood epilepsy. *Brain* 2015;138:2234–2248.
31. Haut SR, Hall CB, Masur J, et al. Seizure occurrence: precipitants and prediction. *Neurology* 2007;69:1905–1910.
32. Scott RC. What are the effects of prolonged seizures in the brain? *Epileptic Disord* 2014;16(Spec No 1):S6–S11.
33. Shegog R, Bamps YA, Patel A, et al. Managing epilepsy well: emerging e-tools for epilepsy self-management. *Epilepsy Behav* 2013;29:133–140.
34. Haut SR. Seizure clustering. *Epilepsy Behav* 2006;8:50–55.
35. Thurman DJ, Beghi E, Begley CE, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia* 2011;52(Suppl 7):2–26.
36. Noble AJ, Marson AG. Which outcomes should we measure in adult epilepsy trials? The views of people with epilepsy and informal carers. *Epilepsy Behav* 2016;59:105–110.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1. Comparison of seizure events for circadian hours with the reference hour (hour 01:00 a.m.) in a mixed-effects negative binomial model of 9,849 patients.

Table S2. Comparison of seizure events for days of the week with the reference day (Sunday) in a mixed-effects zero-inflated negative binomial model of 9,849 patients.

Figure S1. Comparison of distribution of patient-reported monthly seizure frequencies in adults (red) and children (blue) with epilepsy.

Figure S2. (A) Histograms of the log interseizure interval (ISI) distribution. The distributions are highly symmetric, with the skewness of the distribution of log interseizure intervals being −0.26 (slightly skewed left) for adults and 0.08 (slightly skewed right) for children. (B) Histogram of all interseizure intervals distribution. (C) Histogram of median interseizure interval (for each patient) distribution.