University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Karoly, PJ;Romero, J;Cook, MJ;Freestone, DR;Goldenholz, DM

Title:
When can we trust responders? Serious concerns when using 50% response rate to assess clinical trials

Date:
2019-09-01

Citation:
Karoly, P. J., Romero, J., Cook, M. J., Freestone, D. R. & Goldenholz, D. M. (2019). When can we trust responders? Serious concerns when using 50% response rate to assess clinical trials. Epilepsia, 60 (9), pp.E99-E103. https://doi.org/10.1111/epi.16321.

Persistent Link:
https://hdl.handle.net/11343/286348

Check for updates

DR. PHILIPPA J KAROLY (Orcid ID : 0000-0002-9879-5854)
DR. DANIEL M GOLDENHOLZ (Orcid ID : 0000-0002-8370-2758)

Author Manuscript

Article type : Brief Communication (includes Case Reports)

Title: When can we trust responders? Serious concerns when using 50% response rate to assess clinical trials

Running Title: When can we trust responders?

Authors: Philippa J. Karoly[1], Juan Romero[4], Mark J. Cook[1,2], Dean R. Freestone[3], Daniel M. Goldenholz[4]

1. Graeme Clark Institute, The University of Melbourne, VIC Australia
2. St Vincent's Hospital, Department of Medicine, The University of Melbourne, VIC Australia
3. Seer Medical, VIC Australia
4. Beth Israel Deaconess Medical Center, Harvard Medical School, MA USA

Corresponding author: Philippa J. Karoly, karolyp@unimelb.edu.au, Graeme Clark Institute, 203 Bouverie Street Carlton VIC 3053

Number of pages: 6
Number of words: 1806
Number of figures: 1
Number of tables: 1

Summary: Individuals' seizure rates are highly volatile, with large fluctuations from month-to-month. Nevertheless, changes in individuals' mean seizure rates are used to measure whether or not trial participants successfully respond to treatment. This study aims to

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/EPI.16321

This article is protected by copyright. All rights reserved

quantify the challenges in identifying individual treatment responders in epilepsy. A power calculation was performed to determine the trial duration required to detect a significant 50% decrease in seizure rates ($p < 0.05$) for individuals. Seizure rate simulations were also performed to determine the number of people who would appear to be 50% responders by chance. Seizure rate statistics were derived from long-term seizure counts recorded during a previous clinical trial for an implantable seizure monitoring device. We showed that individual variance in monthly seizure rates can lead to an unacceptably high false positive rate in the detection of individual treatment responders. This error rate cannot be reduced by increasing the trial population; however, it can be reduced by increasing the duration of clinical trials. This finding suggests some drugs may be incorrectly evaluated as effective; or, conversely, helpful drugs could be rejected based on 50% response rates. It is important to pursue more nuanced approaches to measuring individual's treatment response, which consider the patient-specific distributions of seizure rates.

Key words: epilepsy, seizure rate, treatment, clinical trials, response rate

Author Manuscript

INTRODUCTION

Seizure rates show large fluctuations from month to month. Previous work has demonstrated that a memoryless statistical distribution can reasonably approximate seizure diaries[1]. In statistical terms, the memoryless distribution of most people's seizure counts have been shown to be overdispersed compared to a Poisson distribution, (i.e. the variance is higher than the mean)[2,3]. This overdispersion is observed across diverse population studies and does not depend on the underlying rate of seizures[2]. In fact, there is a power-law relationship between the variance and the mean seizure rate[2]. This relationship shows that fluctuations in seizure rate scale in accordance with mean seizure rate, so regardless of whether someone has frequent or infrequent seizures they will see large deviations above and below their baseline seizure rate. Consequently, the mean seizure rate alone does not provide enough information to determine whether short-term changes are statistically meaningful. The distribution, or at least the variance, of someone's seizure rates are required to assess if seizure rates have really increased or decreased. Nevertheless, changes in individual seizure rates are overwhelmingly used to assess drug efficacy in clinical trials. The percentage of "responders" (patients with at least a 50% reduction in seizure rates, RR50) remains a primary endpoint for clinical trials.

This article is protected by copyright. All rights reserved

This study demonstrates that there may be unacceptably high error rates ($p > 0.05$) when detecting individual responders.

METHODS

We used data from a previous trial for an implantable seizure warning device (the "NeuroVista trial")[4] to determine expected seizure rates and standard deviations. The total clinical seizure count per month was computed for 15 subjects. Recording durations ranged from 6-months to 2-years and subjects were undergoing their usual day-to-day activities, including taking anti-seizure medication. For further details on subject demographics the reader is referred to Cook et al (2013). It should be noted that subjects all had refractory seizures, and most were taking multiple medications (see Table 1). Subjects' medications remained stable throughout the recording period, although in the last month of recording there was the addition of clonazepam during predicted high risk periods. Hence seizure numbers from the final month were excluded from this analysis. One subject (S1) was excluded from this study as their medication regime changed halfway through the recording period. All seizures were recorded on intracranial EEG and reviewed by expert markers. Seizure symptoms were either self-reported, reported by carers or confirmed via audio recording. Seizures without confirmed symptoms that appeared electrographically identical to clinical seizures were also included. The present study only made use of the redacted seizure diaries produced from the above EEG analysis.

A standard 80% power calculation was used to determine how many samples (months) were needed to detect a 50% decrease in seizure rates. Sample size, N, is given as[5]:

$$N = \frac{2\sigma^2(Z_\alpha + Z_{(1-\beta)})^2}{\Delta^2} \qquad (1)$$

where Z_α and $Z_{1-\beta}$ are constants based on the desired Type I (α) and Type II (β) errors, Δ is the desired effect size (the difference in seizure rate) and σ^2 is the sample variance of mean monthly seizure rates. Power calculations typically use $\alpha = 0.05$ and $\beta = 0.20$, which allows for a false positive rate of 5% and a false negative rate of 20%. Based on these values, we used $Z_\alpha = 1.65$ (for a one-sided significance test) and $Z_{1-\beta} = 0.8416$. Variance, σ^2, was calculated from the monthly seizure rates recorded during the NeuroVista trial. The mean

This article is protected by copyright. All rights reserved

Author Manuscript

monthly seizure rate, μ, was also computed and the effect size, Δ, was set to $\frac{\mu}{2}$ seizures to simulate a 50% reduction in mean seizure rate.

Seizure rate simulations were performed to determine the expected false positive rate of the 50% response rate, that is the number of people who would appear to be 50% responders by chance. Every simulation used 1000 participants, where a given participant was randomly assigned a mean and standard deviation corresponding to the actual mean and standard deviation of monthly seizure rates recorded from a randomly chosen subject in the NeuroVista trial. NeuroVista subjects with a mean seizure rate of less than 2 seizures per month, or with medication changes (subject 1), were not included in simulations. For each simulated patient, a random sequence of monthly seizure rates was generated by sampling from a normal distribution and rounding to the nearest whole seizure count. The negative binomial distribution was also used for simulations, with parameters estimated from individuals' monthly seizure counts using a maximum likelihood estimation. Where randomly generated seizure rates were negative, the rate was set to zero. For each simulated patient, the percentage change in seizure rates was computed as the difference in mean monthly seizure frequency between a baseline period and treatment period. The duration of the baseline and treatment periods were altered to simulate different trial parameters. The total simulation duration (months) was the sum of the baseline and treatment periods. Simulations were generated for different trial durations, where the treatment duration ranged from 1 to 12 months (step size = 1). The baseline duration was set to either a fixed 1- or 2-month duration, or equal to the treatment duration (1 – 12 months). For each trial duration, the simulation was repeated 1000 times (i.e. in total we simulated 1000 trials each with 1000 patients resulting in analysis of 1,000,000 individuals).

Note that simulations were not designed to imitate clinical trials in the sense that no intervention was actually applied in the treatment period – representing the idealized placebo situation. For each subject, the baseline and treatment periods were sampled from an identical distribution. In this way, simulations were designed to explore fluctuations in seizure rates that occur by chance in a population of people with epilepsy.

RESULTS

This article is protected by copyright. All rights reserved

Author Manuscript

Using historical seizure records of individual mean seizure rates and standard deviations, we performed a sample size calculation to determine how long it would take to reliably detect whether an individual was a 50%-responder with 80% power. Table 1 shows the mean and standard deviation of monthly seizure rates that were reported for the duration of the NeuroVista trial, as well as the required sample size to achieve 80% power (given by Eqn. 1). Here, sample size can be interpreted as the number of months required to detect a significant 50% decrease in an individual's mean seizure rate. Subjects typically required over 6 months (median of 20 months) to determine whether a 50% decrease was significant ($p < 0.05$).

Using the Kolmogorov-Smirnov test the null hypothesis that the data were normally distributed was rejected for two out of the fifteen patients (Patient 12, $p < 0.05$ and Patient 14, $p < 0.01$). These subjects had the lowest seizure rates (mean of less than one seizure per month) and were not used in simulations. Such patients would typically be excluded by standard inclusion criteria of epilepsy trials; therefore, the use of the normal distribution for seizure rate simulations was justified. An autocorrelation analysis (with lags from one to 20 months) was performed to determine whether individuals' monthly seizure rates were uncorrelated from month to month or showed some dependence on previous data points. One individual (S6) showed significant autocorrelation at a lag of two months ($p < 0.05$). No other patients showed significant autocorrelation, providing some support for the assumption of independent sampling used in simulations. This aligns with our previous work showing that memoryless distributions are approximately equivalent to those that include hidden Markovian states representing memory of prior days' seizures[1].

We used simulations to explore how individual variability impacts the detection of 50%-responders at the population level. Simulations computed the expected false positive rates of 50%-responders in a clinical trial population (i.e. the rate of people who appear to be responders by chance alone) across different trial baseline and treatment periods. Figure 1 shows the expected false positive rate based on a normal and negative binomial distribution of monthly seizure rates. It can be seen from Figure 1A that for equal baseline and treatment periods, the false positive rate dropped below 5% after a treatment phase of approximately four months (i.e. a total trial duration of 4-month baseline + 4-month treatment = 8 months). However, when only two months were used to measure the baseline seizure rate, the response rate did not reach significance until seven months of treatment (total trial duration of nine

This article is protected by copyright. All rights reserved

Author Manuscript

months). The trial with two-month baseline did not achieve a false positive rate of less than 5% under the assumption of a negative binomial distribution of seizure rates (Figure 1B).

DISCUSSION

Individual variance in monthly seizure rates means it may take a long time to reliably detect a 50% decrease in seizure rates (Table 1). At the population level this can lead to an unacceptably high false positive rate ($p > 0.05$, i.e. more than 1 in 20) in the detection of individual treatment responders (Figure 1). The high rate of false positive responders using a 2-month baseline period was especially concerning, as a baseline of 8-weeks is typical for trials[6]. This error rate cannot be reduced by increasing the trial population (our simulations used 1000 patients); however, it can be reduced by increasing the duration of clinical trials.

The 50% response rate remains the preferred outcome measure of the European Medicines Agency[7]. The U.S. Food and Drug Administration prefers median-%-reduction (MPR). The MPR is more statistically efficient than the 50% response rate[8], which is likely due to the median's resistance to large fluctuations. However, there is no widely accepted method for computing a sample size based on a difference in medians[9]. When sample size calculation is based on the 50% responder rate then it is important to consider how the population variance is estimated[5]. The individual variance in seizure rates is likely to translate to a high population variance governing the outcome of clinical trials that assess differences in responder proportions between trial arms. Future work will focus on precisely quantifying the impact of individual variance on group level statistics used in clinical trials.

It is important to note that the reported false positive rates are based on the subjects in the NeuroVista trial, and do not represent the full range of participants enrolled in clinical trials. The NeuroVista cohort all had medically refractory, focal epilepsy which may manifest with different seizure rates compared to other epilepsy syndromes. However, we note that earlier work demonstrated that the seizure rate mean and variance for the NeuroVista participants was consistent with self-reported diary data from a large cohort[2]. The assumption that monthly seizure rates are independently distributed is likely to be violated by some individuals at particular time scales, especially considering the prevalence of slow, multiday cyclic trends[10]. In this study, one individual showed significant correlation in their monthly seizure rates at a lag of two months. This study also did not record drug effects (i.e. missed medication), which could affect seizure rate variance. During clinical trials the variance may

This article is protected by copyright. All rights reserved

Author Manuscript

be somewhat reduced by controlling this environmental variability; however, it is worth nothing that in a trial patient self-reporting errors may also decrease the signal-to-noise ratio compared to intracranial monitoring.

This study highlights how the low signal-to-noise ratio of individual seizure rates can make it challenging to assess treatment efficacy within a population. The traditional approach to this problem is to sample a large number of patients because it is only a group comparison (drug versus placebo) that is ultimately of interest for the "success" of a trial. However, population response rates are difficult to assess when there is higher uncertainty about whether individuals truly responded. Due to the limitations of this study, it is difficult to definitively state how individual seizure rate fluctuations alter group level 50% response rates. The presented results should not be taken to invalidate the success or failure of existing clinical trials. However, the assumptions used to calculate trial power and significance should be evaluated in light of the expected high false positive rate of 50% responders. Continually assessing and improving clinical trials is important considering anti-epileptic drugs showed modest efficacy across a meta-analysis of clinical trials[11], and for decades there has been no improvement in the number of people with medically refractory epilepsy[12]. Developing more nuanced evaluation metrics may help to detect a treatment response in this challenging cohort of people with refractory seizures.

In conclusion, it is difficult to detect a true difference between individual treatment responders and non-responders, particularly when the trial duration is short (i.e. < 8 months). This challenge may result in drugs being incorrectly evaluated as effective; or, conversely, could cause helpful drugs to be rejected, due to potentially high rates of false positive responders in the treatment or control arm of a clinical trial. To address this challenge, it is important to pursue and validate novel metrics for computing drug efficacy that consider patient-specific models of seizure rates. For example, models that account for the individual variance of seizure rates[13] may provide a better chance of distinguishing responders from non-responders. We anticipate that digital seizure diaries and wearable devices could one day enable *a-priori* models of individual's seizure patterns, providing valuable pre-trial screening tools.

ACKNOWLEDGMENTS

This research was supported by a National Health and Medical Research Council Grant,

This article is protected by copyright. All rights reserved

Author Manuscript

Australia (NHMRC Project APP1065638).

CONFLICTS OF INTEREST

None of the authors has any conflict 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

REFERENCES

1. 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(5):835–844.

2. Goldenholz DM, Goldenholz SR, Moss R, et al. Is seizure frequency variance a predictable quantity? Annals of Clinical and Translational Neurology. 2017; 5(2):201–7.

3. Karoly PJ, Goldenholz DM, Cook MJ. Are the days of counting seizures numbered? Current Opinion in Neurology. 2018; 31(2):162–8.

4. Cook MJ, O'Brien TJ, Berkovic SF, 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. The Lancet Neurology. 2013; 12(6):563–71.

5. Cohen J. Statistical power analysis for the social sciences. 2nd ed. USA: Lawrence Erlbaum Associates; 1988.

6. 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):124–131.

7. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders [Internet]. European Medicines Agency; 2010 [cited 2017]. Available from: http://www.ema.europa.eu

8. Siddiqui O, Hershkowitz N. Primary Efficacy Endpoint in Clinical Trials of Antiepileptic Drugs: Change or Percentage Change. Drug Information Journal. 2010; 44(3):343–50.

Author Manuscript

This article is protected by copyright. All rights reserved

9. O'Keeffe AG, Ambler G, Barber JA. Sample size calculations based on a difference in medians for positively skewed outcomes in health care studies. BMC Med Res Methodol [Internet]. 2017; 17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5712177/

10. Karoly PJ, Goldenholz DM, Freestone DR, et al. Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. The Lancet Neurology. 2018; 17(11):977–85.

11. Beyenburg S, Stavem K, Schmidt D. Placebo-corrected efficacy of modern antiepileptic drugs for refractory epilepsy: systematic review and meta-analysis. Epilepsia. 2010; 51(1):7–26.

12. Chen Z, Brodie MJ, Liew D, et al. Treatment outcomes in patients with newly diagnosed epilepsy treated with established and new antiepileptic drugs: a 30-year longitudinal cohort study. JAMA neurology. 2018; 75(3):279–286.

13. Goldenholz DM, Goldenholz SR, Moss R, et al. Does accounting for seizure frequency variability increase clinical trial power? Epilepsy Research. 2017; 137(Supplement C):145–51.

Author Manuscript

FIGURE CAPTIONS

Figure 1: False positive rate when detecting 50% responders. Simulations were performed to determine the proportion of people who appeared to be 50%-responders in a population of 1000 people with epilepsy. The mean false positive rate across simulations is shown on the y-axis. The shaded region shows the 95% confidence intervals using the standard deviation. Simulations were performed 1000 times for each treatment period (x-axis). Different baseline periods were also considered (a 1-month and 2-month baseline monitoring period, and a baseline the same duration as the treatment period). A) Simulations based on a normal distribution, using the mean and standard deviation of seizure rates observed in the NeuroVista cohort (Table 1). B) Simulations based on a negative binomial distribution with parameters estimated from the monthly seizure count data recorded in the NeuroVista cohort.

This article is protected by copyright. All rights reserved

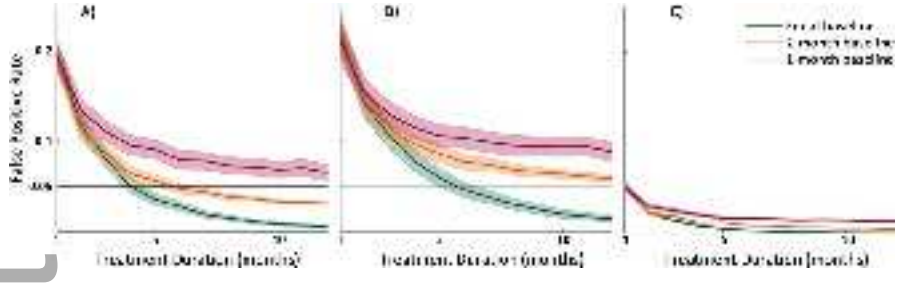
TABLES

Table 1: Seizure rate statistics and 50% response power. Recording durations, medications and monthly seizure rate means and standard deviations for subjects in the NeuroVista trial. N is the number of months required to assess whether a person showed >50% decrease in seizure rates ($p < 0.05$) based on a sample size calculation with 80% power.

Subject	*Recording (months)*	*Anti-seizure medication*	*Mean*	*Standard deviation*	*N (Months)*
1	27	CLZ, LEV, LTG, VPA	5.7	2.9	13
2	23	LCM, LTG, OXC, VPA	1.4	1.0	29
3	20	CBZ, LTG, PHT	19.1	12.8	22
4	8	CBZ, LCM, LTG, TPM, PHT	2.7	0.7	4
5	8	CLZ, LTG, OXC, TPM	1.1	0.9	30
6	12	None	6.0	2.6	9
7	7	CBZ, CLZ, LEV	54.3	53.6	48
8	20	CBZ, LEV	23.8	15.5	21
9	14	CBZ	14.3	5.4	7
10	14	LEV, OXC, ZNS	40.8	26.3	20
11	24	LCM, PHT, PRP	20.3	12.2	18
12	22	LTG, LCM, PHT, RTG	0.6	1.3	239
13	24	CBZ, CLZ, LEV, LCM	21.2	11.9	16
14	19	CLZ, OXC	0.6	1.0	127
15	15	CBZ, LCM, PRP, TPM	5.1	2.0	7

Author Manuscript

This article is protected by copyright. All rights reserved



epi_16321_f1.tif

Author Manuscript

This article is protected by copyright. All rights reserved