Epilepsy Research 137 (2017) 145–151

Contents lists available at ScienceDirect

Epilepsy Research

journal homepage: www.elsevier.com/locate/epilepsyres

Does accounting for seizure frequency variability increase clinical trial power?

Daniel M. Goldenholz[a,j,*], Shira R. Goldenholz[j], Robert Moss[b], Jacqueline French[c], Daniel Lowenstein[d], Ruben Kuzniecky[c], Sheryl Haut[e], Sabrina Cristofaro[c], Kamil Detyniecki[f], John Hixson[d], Philippa Karoly[g], Mark Cook[g], Alex Strashny[h], William H Theodore[a], Carl Pieper[i]

[a] Clinical Epilepsy Section, NINDS, NIH, United States
[b] SeizureTracker LLC, United States
[c] Department of Neurology,New York University, United States
[d] Department of Neurology, UCSF, United States
[e] Department of Neurology, Montefiore Medical Center/Albert Einstein College of Medicine, United States
[f] Department of Neurology, Yale University, United States
[g] University of Melbourne, Australia
[h] Department of Neurology, Centers for Disease Control, United States
[i] Duke University Medical Center, Dept. of Biostatistics and Bioinformatics, United States
[j] Division of Epilepsy, Beth Israel Deaconess Medical Center

ARTICLE INFO

Keywords:
Epilepsy
Clinical trials
Placebo effect
Prediction
Seizure frequency
Natural variability

ABSTRACT

Objective: Seizure frequency variability is associated with placebo responses in randomized controlled trials (RCT). Increased variability can result in drug misclassification and, hence, decreased statistical power. We investigated a new method that directly incorporated variability into RCT analysis, Z_V.

Methods: Two models were assessed: the traditional 50%-responder rate (RR50), and the variability-corrected score, Z_V. Each predicted seizure frequency upper and lower limits using prior seizures. Accuracy was defined as percentage of time-intervals when the observed seizure frequencies were within the predicted limits. First, we tested the Z_V method on three datasets (SeizureTracker: n = 3016, Human Epilepsy Project: n = 107, and NeuroVista: n = 15). An additional independent SeizureTracker validation dataset was used to generate a set of 200 simulated trials each for 5 different sample sizes (total N = 100 to 500 by 100), assuming 20% dropout and 30% drug efficacy. "Power" was determined as the percentage of trials successfully distinguishing placebo from drug ($p < 0.05$).

Results: Prediction accuracy across datasets was, Z_V: 91–100%, RR50: 42–80%. Simulated RCT Z_V analysis achieved > 90% power at N = 100 per arm while RR50 required N = 200 per arm.

Significance: Z_V may increase the statistical power of an RCT relative to the traditional RR50.

1. Introduction

There is a need for new epilepsy drugs, given the 35% prevalence of drug-resistant epilepsy (Brodie et al., 2012; Kwan and Brodie, 2000). However, drug development remains challenging due to high expense and frequent trial failure. Trials suffered from rising placebo response rates over the past several decades (Rheims et al., 2011), typically 4–27% (Goldenholz and Goldenholz, 2016) but recently up to 40% (French et al., 2015). This can translate into unsuccessful trials (Halford et al., 2011), increased sample size, and increased development costs (PhRMA, 2015). Seizure frequency variability at the patient level, typically unreported, may explain a significant portion of placebo responses, because natural frequency fluctuations are sufficiently large to produce a "response" even without treatment (Goldenholz et al., 2015). Uncertainty about variability may hamper randomized clinical trial (RCT) interpretation. With current methods, variability represents "noise" obscuring the drug efficacy "signal". With lower noise, trials are expected to cost less and have fewer failures.

* Corresponding author at: Beth Israel Deaconess Medical Center, 330 Brookline Ave, Baker 5, Boston MA 02115, United States.
E-mail addresses: daniel.goldenholz@bidmc.harvard.edu (D.M. Goldenholz), shira.r.g@gmail.com (S.R. Goldenholz), rob@seizuretracker.com (R. Moss), jacqueline.french@nyumc.org (J. French), Lowenstein@ucsf.edu (D. Lowenstein), Ruben.Kuzniecky@nyumc.org (R. Kuzniecky), SHAUT@montefiore.org (S. Haut), Sabrina.Cristofaro@nyumc.org (S. Cristofaro), kamil.detyniecki@yale.edu (K. Detyniecki), John.Hixson@ucsf.edu (J. Hixson), p.karoly@student.unimelb.edu.au (P. Karoly), markcook@unimelb.edu.au (M. Cook), kpr9@cdc.gov (A. Strashny), TheodorW@ninds.nih.gov (W.H. Theodore), carl.pieper@duke.edu (C. Pieper).

http://dx.doi.org/10.1016/j.eplepsyres.2017.07.013
Received 3 May 2017; Received in revised form 28 June 2017; Accepted 21 July 2017
Available online 25 July 2017
0920-1211/ Published by Elsevier B.V.

The RR50 (the percentage of patients with 50% seizure reduction in each trial arm), is the preferred outcome measure of the European Medicines Agency (EMA) (European Medical Agencies, 2010). The U.S. Food and Drug Administration prefers median-%-change (MPC). Trials typically require co-primary RR50 and MPC endpoints. RR50 is less statistically efficient than MPC (Siddiqui and Hershkowitz, 2010), and typically used in power calculations for patient enrollment. However, based on recent evidence, the RR50 likely overestimates clinically relevant measures (Goldenholz et al., 2015). Simulations based on 1767 patient seizure diaries show that many RCT 50%-responders may subsequently become non-responders due to large natural variability. Consequently, models incorporating expected variability may improve epilepsy RCT interpretability, generalizability, and efficiency. Obviously, such models would only be of use if adopted by regulatory agencies.

Standard clinical practice includes implicit judgments about natural variability as well. Physicians are expected to make medication changes based on whether seizure rates have exceeded some arbitrary upper bound. If a drug adjustment results in rate decreases below an arbitrary lower boundary, the adjustment is considered beneficial. For patients with years of seizure-freedom, variability computations are irrelevant. But if seizure-freedom is short-lived, measured over a short duration, or if the patient is not seizure-free, no formal clinical tools exist to calculate expected bounds on seizure rates.

Clinicians and trialists would benefit from a robust method for predicting natural seizure frequency variability. This study represents the first attempt to account for the impact of variability on seizure frequency measurements, using a multi-modal data-driven approach.

2. Materials and methods

2.1. Overview

This work presents a novel method for assessing RCTs called Z_V (Methods 2.2). Z_V and RR50 were compared in their ability to predict seizure frequencies several months into the future (Fig. 1, Methods 2.3). In three datasets, each patient diary was divided into 6-month intervals to mimic typical RCT duration (Perucca, 2012). In each interval, early seizure rates were used to predict later rates using RR50 and Z_V.

To assess Z_V utility in an RCT (Fig. 2), we generated a set of simulated clinical trials based on realistic seizure data (Methods 2.4). Five sets of 200 trials each included 100, 200, 300, 400 or 500 patients.

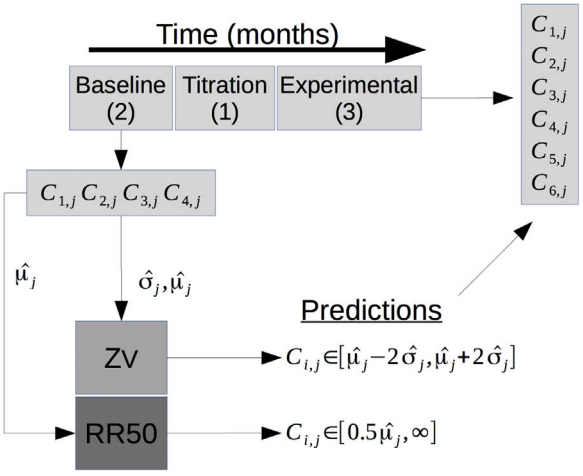

Fig. 1. Prediction models. The three phases of a clinical trial are shown: baseline (B), titration (T) and experimental (E). Placebo is given during T and E for those patients assigned to placebo. Drug is titrated up during T, and given at a steady dose during E. The 2 prediction models use the measured $\hat{\mu}_j$ and $\hat{\sigma}_j$ (the mean and standard deviation of 2-week seizure counts) from the baseline period, to predict the limits of $C_{i,j}$, the 2-week seizure counts during the experimental phase.

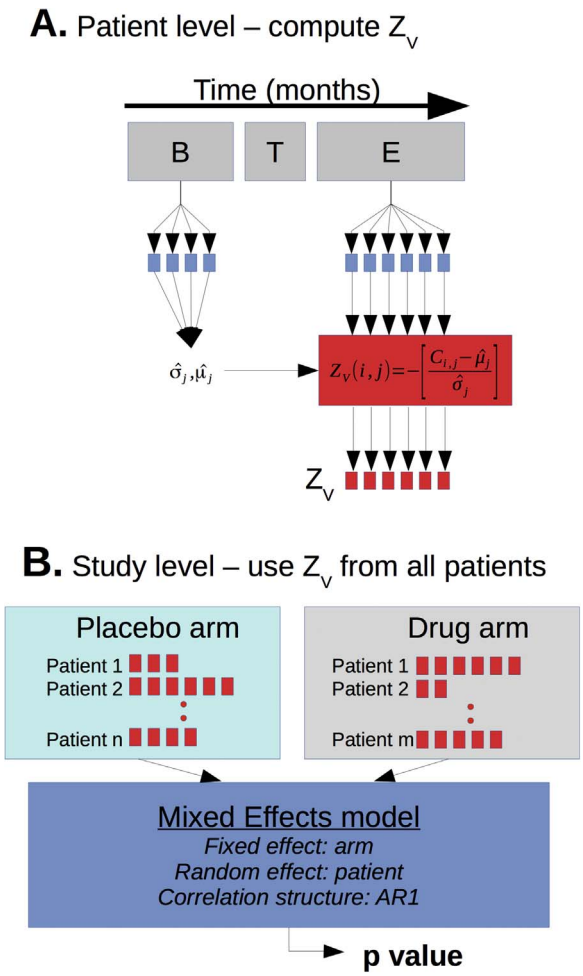

Fig. 2. The Z_V analysis method. A. Z_V is calculated for a single patient. A typical trial constructed with baseline (B), titration (T) and experimental (E) phases is shown. The baseline is divided 4 segments in this image, however this number is flexible. Those segments are used to calculate measured $\mu_{Baseline}$ and $\sigma_{Baseline}$, the mean and standard deviation of the seizure counts from each segment. These are then used to compute normalized Z_V from the similarly divided segments of E. Note that in this Image 6 segments are represented, though this number is flexible. B. At the study level, all patients contribute a set of Z_V values, however if a patient drops out early then they may contribute less than a full set. Dropout is represented when not all 6 Z_V squares are present for each patient. All completed Z_V values from each arm are treated with equal weight and are compared with a mixed effects model to obtain a final p value.

Statistical power was computed for each series, and each calculation method (RR50, MPC and Z_V), to determine the minimum number of patients needed for the trial to achieve 90% power for each method.

2.2. The variability-corrected Z_V method

The Z_V model assumed seizure frequency variability during both experimental and baseline periods remained unchanged. Typically, "seizure frequency" refers to a 28-day seizure count; here, we focus on 14-day seizure counts. For mathematical simplicity, an individual's seizure frequencies were assumed to follow a Gaussian distribution. Each patient's seizure count for each 2-week interval of time was represented by $C_{i,j}$, the count of the i^{th} interval in the j^{th} patient. Because we chose a 2-month baseline (Fig. 1), 4 intervals of 2-weeks were considered. The model calculated an estimated mean ($\hat{\mu}_j$), and standard deviation ($\hat{\sigma}_j$) of the set of $C_{i,j}$'s during the baseline (Eqs. (1),(2)), with the 4 values of $C_{i,j}$ (M = 4):