CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Treatment trials for neonatal seizures: the effect of design on sample size
Authors	Stevenson, Nathan J.;Boylan, Geraldine B.;Hellström-Westas, Lena;Vanhatalo, Sampsa
Publication date	2016-11-08
Original Citation	Stevenson, N. J., Boylan, G. B., Hellström-Westas, L. and Vanhatalo, S. (2016) 'Treatment trials for neonatal seizures: the effect of design on sample size', PLoS ONE, 11(11), e0165693 (14pp). doi:10.1371/journal.pone.0165693
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1371/journal.pone.0165693
Rights	© 2016, Stevenson et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-30 20:49:37
Item downloaded from	https://hdl.handle.net/10468/3367

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

PLOS ONE

RESEARCH ARTICLE

Treatment Trials for Neonatal Seizures: The Effect of Design on Sample Size

Nathan J. Stevenson[1,4]*, Geraldine B. Boylan[1], Lena Hellström-Westas[2], Sampsa Vanhatalo[3,4]

1 Irish Centre for Fetal and Neonatal Translational Research and Department of Paediatrics and Child Health, University College Cork, Cork, Ireland, **2** Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden, **3** Department of Clinical Neurophysiology HUS Medical Imaging Center, Helsinki University Central Hospital, Helsinki, Finland, **4** Department of Neurological Sciences, Clinicum, University of Helsinki, Helsinki, Finland

* nathan.stevenson@helsinki.fi

CrossMark
click for updates

OPEN ACCESS

Citation: Stevenson NJ, Boylan GB, Hellström-Westas L, Vanhatalo S (2016) Treatment Trials for Neonatal Seizures: The Effect of Design on Sample Size. PLoS ONE 11(11): e0165693. doi:10.1371/journal.pone.0165693

Editor: Giuseppe Biagini, University of Modena and Reggio Emilia, ITALY

Received: June 6, 2016

Accepted: October 17, 2016

Published: November 8, 2016

Copyright: © 2016 Stevenson et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are from two studies on neonatal seizures whose authors may be contacted at the Irish Centre for Fetal and Neonatal Translational Research. These studies relate to projects funded by the Wellcome Trust UK (85249/z/08/z) and Health Research Board, Ireland (RP/2008/238). The original publications based on these data are listed in the manuscript.

Funding: NJS was supported by an EU Marie Skłodowska Curie Individual Fellowship (H2020-MCSA-IF-656131), GBB was supported by Science Foundation Ireland (12/RC/2272). SV was supported by Academy of Finland (#276523 and

Abstract

Neonatal seizures are common in the neonatal intensive care unit. Clinicians treat these seizures with several anti-epileptic drugs (AEDs) to reduce seizures in a neonate. Current AEDs exhibit sub-optimal efficacy and several randomized control trials (RCT) of novel AEDs are planned. The aim of this study was to measure the influence of trial design on the required sample size of a RCT. We used seizure time courses from 41 term neonates with hypoxic ischaemic encephalopathy to build seizure treatment trial simulations. We used five outcome measures, three AED protocols, eight treatment delays from seizure onset (T_d) and four levels of trial AED efficacy to simulate different RCTs. We performed power calculations for each RCT design and analysed the resultant sample size. We also assessed the rate of false positives, or placebo effect, in typical uncontrolled studies. We found that the false positive rate ranged from 5 to 85% of patients depending on RCT design. For controlled trials, the choice of outcome measure had the largest effect on sample size with median differences of 30.7 fold (IQR: 13.7–40.0) across a range of AED protocols, T_d and trial AED efficacy ($p<0.001$). RCTs that compared the trial AED with positive controls required sample sizes with a median fold increase of 3.2 (IQR: 1.9–11.9; $p<0.001$). Delays in AED administration from seizure onset also increased the required sample size 2.1 fold (IQR: 1.7–2.9; $p<0.001$). Subgroup analysis showed that RCTs in neonates treated with hypothermia required a median fold increase in sample size of 2.6 (IQR: 2.4–3.0) compared to trials in normothermic neonates ($p<0.001$). These results show that RCT design has a profound influence on the required sample size. Trials that use a control group, appropriate outcome measure, and control for differences in T_d between groups in analysis will be valid and minimise sample size.

#288220), Sigrid Juselius Foundation, and the European Union (FP7-HEALTH-2009-4.2-1, grant agreement 241479). The data used in this study was supported by Principal Investigator Awards to GBB from the Health Research Board (RP/2008/238) and the Wellcome Trust (085249). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Evidence-based guidelines of drug treatments require studies where the drug effect is quantitatively measured and statistically compared to an alternative treatment or placebo. Key elements of successful trial designs include the choice of relevant outcome measure(s) and sample size. Selection of the outcome measure is typically a compromise between what is known to be important in pathophysiology and what is practically possible. While the sample size must be selected so that it is large enough to demonstrate statistical significance of a clinically relevant effect, there is a practical need to minimize the sample size via trial design in order to reduce the cost and duration of trials in vulnerable patient groups with rare conditions. The details of RCT design are, therefore, critical when interpreting study findings; particularly when study findings lead to a change in clinical practice.

Measuring treatment outcome is challenging when the natural course of the illness is variable and has a tendency to improve [1, 2]. It becomes even more difficult when the natural duration of the illness is short relative to the timing of treatment protocols [3]. Many neurological illnesses, such as status epilepticus, and migraine, are well known to have highly variable and self-limiting time courses. This is further complicated by the fact that the partial effectiveness of existing medications (such as phenobarbitone) may preclude the use of a placebo resulting instead in the use of a positive control group [4, 5].

Seizures in neonates present a particular challenge in this context because seizures tend to resolve within tens of hours, leaving little time to observe treatment effects [6, 7]. In addition, seizure occurrence is highly variable across neonates and varies with aetiology [7–10]. It is generally accepted that the accumulated duration of seizures, or seizure burden, as measured by multi-channel EEG, should be the quantitative measure of choice when assessing anti-epileptic drug (AED) efficacy [4, 5, 11–13]. This is based on the assumption that a high seizure burden causes further damage to the already compromised neonatal brain [14–16]. An evidence based AED trial should, therefore, be powered so that it takes into account the natural time course of neonatal seizures. Otherwise, a bias in outcome towards a positive treatment effect could be introduced. It has been difficult to accurately power a neonatal AED trial *a priori*, due to limited data on the temporal behaviour of seizures over a period of days. Recent advances in long term EEG monitoring have shed light onto the temporal evolution of neonatal seizures which will improve power calculations [7, 10].

In the present work, we aimed to establish the effect of different designs of neonatal AED trials on the sample size estimate from a power analysis. To this end, we modelled seizure time courses based on real data, which allowed us to examine the influence of trial design on the sample size for different outcome measures, AED protocol and delays in intervention at different levels of AED efficacy.

Methods

This study was performed as a series of simulated AED trials for neonatal seizures which allows direct comparison between different trial designs. The key prior knowledge required for realistic simulations is the natural time course of seizure burden during neonatal seizures. Here, we took advantage of two recently collected cohorts of long-term EEG recordings from neonates with seizures due to hypoxic ischemic encephalopathy (HIE) [7, 10]. These cohorts contained complete, second by second, quantification of seizure burden in two cohorts of neonates collected before (n = 18, normothermic group) and after (n = 23) the introduction of hypothermia as treatment for HIE [7, 10]. In these studies, data collection was conducted with approval from the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Ireland. Parental written, informed consent was obtained for all newborns recruited for EEG monitoring studies.

All data were anonymised. For more details on the demographics and seizure burden of this cohort see the S1 Appendix and Lynch et al. (2012 and 2014) [7, 10]. The most significant difference between these two cohorts was a lower overall seizure burden in neonates treated with hypothermia.

Generation of a realistic model of neonatal seizure burden time courses

We first constructed a realistic model of seizure burden time courses that permitted the use of large simulated cohorts. A lognormal function was found to adequately simulate the seizure burden time course with the characteristic positive skewness that is seen in the distribution of seizures over time. In other words, neonatal seizure burden has been shown to accumulate rapidly in the hours after seizure onset, followed by a more gradual accumulation towards seizure offset [7]. It also results in smoothed time courses where seizures occur to some extent for the entire period of simulation. The smoothing of any discontinuities in the seizure burden time courses was required for the following reasons: 1) discontinuities generated by the response to AED treatment and periods of missing data contaminate the normative seizure burden time course resulting in the need for interpolation, 2) the systematic implementation of the RCT simulations would result in invalid outcome measures in the presence of discontinuity, i.e. no seizure burden (further, an AED would not be given if no seizures were present), 3) the function of seizure burden over time is not a raw value and must be calculated from the raw seizure annotations, so smoothing the seizure burden can be considered as calculating the seizure burden over a longer time period (see S1 Appendix). Notably, while smoothing may reduce the peak seizure burden, it does not significantly alter other important summary measures of seizure burden such as the total seizure burden, the skew towards seizure onset, and the time from seizure onset to the point of maximum seizure burden.

In order to generate a single seizure burden time course, the parameters of the lognormal function were assumed to be random and were selected from a multi-variate distribution estimated from real data. An example of the lognormal function fitted to real data and the lognormal function fitted to all 41 neonates in the cohort are shown in Fig 1A and 1B. More details on the process of simulating seizure burden time courses and the quality of fit of a lognormal function to real data is outlined in S2 Appendix.

Simulation of therapeutic trials

The simulated seizure burden time courses were used in a power calculation for several randomized control trials (RCT). Each RCT was defined using several variables: AED protocol, average time relating to the administration of the intervention with respect to seizure onset (T_d), target level of trial AED efficacy and outcome measure.

Three different AED protocols were used (see Fig 1C): first line AED vs. placebo control, first line AED vs. positive control (assumed to be phenobarbitone), and second line AED vs. placebo control (with phenobarbitone as a first line). The efficacy of phenobarbitone was assumed to be a 75% reduction in seizures for 3h [17]. The T_d variable represents delays in the clinical recognition of seizures and the speed of execution of the trial protocol and was varied from 1h to 8h [18]. We used four AED efficacies from a maximum possible effect (100% reduction in seizures for 72h), to more typical effects (80% reduction for 12h, 80% reduction for 6h, and 50% reduction for 12h) that parallel typical target levels of efficacy [4, 12, 13].

Five outcome measures were estimated from the simulated seizure burden time course (for a graphical representation, see Fig 1C). 1) Total seizure burden: The total accumulated duration of seizures between seizure onset and seizure offset (or monitoring cessation). 2) Post-intervention seizure burden for 1h: The accumulated duration of seizures in a 1h time period

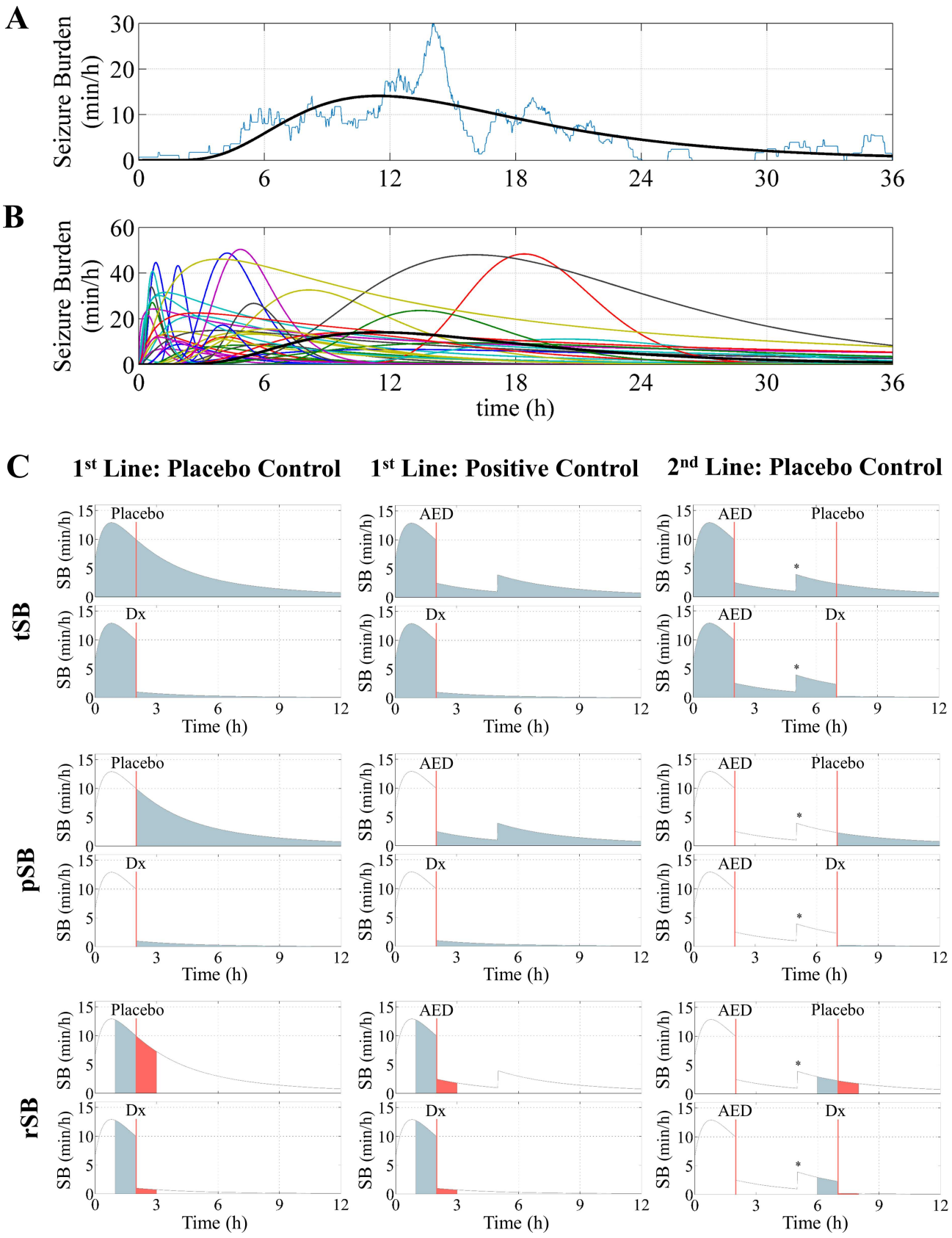

A
Seizure Burden (min/h)
B
Seizure Burden (min/h)
time (h)
C
1st Line: Placebo Control
1st Line: Positive Control
2nd Line: Placebo Control
tSB
pSB
rSB
SB (min/h)
Placebo
AED
Dx
Time (h)

Fig 1. Seizure time courses and neonatal seizure treatment trial designs. The simulated seizure time courses used to estimate the sample size for various designs of a randomized control trial. A) An example smoothed seizure time course from the cohort of Lynch et al. (2015) plotted over the corresponding real seizure time course [10]. B) All 41 smoothed seizure time courses from the cohorts of Lynch et al. (2012 and 2015) showing the variability of seizures in neonates (the black line in A and B refers to the same neonate) [7, 10]. SB/h is seizure burden in minutes per hour; it is a measure of the short term intensity of seizures. Time is measured with respect to seizure onset. C) The use of different AED protocols and outcome measures in RCT design. Each row defines a common outcome measure and each column defines an AED protocol. Outcome measures are defined by the shaded areas: tSB row–total SB (blue shaded area), pSB row–post-intervention SB (blue shaded area), and rSB row–SB response (blue shaded area subtracted from the red shaded area). Treatment delay is the difference between seizure onset and the initiation of the trial protocol and is 2h in these examples. The level of AED efficacy in these examples is an immediate 90% reduction in seizure burden. The trial drug is denoted as Dx. Existing or positive control anti-epileptic drug (AED) effect was based on phenobarbitone: a 75% reduction in seizure burden for 3h. The asterisk denotes the cessation of the existing AED effect (seizure reoccurrence).

doi:10.1371/journal.pone.0165693.g001

after the intervention. 3) Post-intervention seizure burden for 12h: The accumulated duration of seizures in a 12h time period after the intervention. 4) Seizure burden response over 1h: The accumulated duration of seizures in a one-hour time period before the intervention (baseline) subtracted from the accumulated duration of seizures in a 1h time period after the intervention. 5) Seizure burden response over 12h: The accumulated duration of seizures in a one-hour baseline period subtracted from the accumulated duration of seizures averaged across a 12h time period after the intervention.

The total seizure burden is the most general measure of seizures in a neonate and maximising its reduction should have the most positive effect on long term neurodevelopmental outcomes [15]. This outcome measure requires long term, continuous EEG recording from the onset of seizures, which may be logistically challenging in many neonatal intensive care units (NICUs). The post-intervention seizure burden takes into account the seizure burden after the intervention in each patient while the seizure burden before the intervention is ignored. This outcome measure requires a continuous EEG recording from the time of intervention, which is achievable in most NICUs. The seizure burden response is commonly used in uncontrolled studies and requires EEG recording during the pre- and post-intervention time periods.

Several published trials use the seizure burden response as the primary outcome measure [4, 5, 12, 13]. The advantage of the seizure burden response is that it is the only outcome measure that can be used in an uncontrolled study. The total seizure burden has been used to assess treatments such as hypothermia and the usefulness of EEG monitoring to guide treatment [16, 19].

Estimating the sample size (power calculation)

The sample size was calculated using the mean and standard deviations of these outcome measures in simulated arms of the RCT. The sample size was defined as,

$$N = \frac{4(\sigma_1^2 + \sigma_2^2)(1.96 + 0.842)^2}{\Delta^2} \quad (1)$$

where, σ_1 and σ_2 are the standard deviation of the outcome measure in the intervention and control groups, respectively, and Δ is the effect size ($\Delta = \mu_2 - \mu_1$ where μ_1 and μ_2 are the mean of the outcome measure in the trial AED and control group, respectively) [20]. The constants relate to the pre-selection of 80% power and a level of significance of 0.05 (two-sided test). An equal number of patients in each group ($N/2$) was also assumed as per current practice in studies on treatments for neonatal seizures [4, 5, 13, 15, 16].

Simulations and Analysis

A diagrammatic summary of the process of estimating sample size from simulated RCT is shown in Fig 2. In order to calculate the sample size, the mean and standard deviation of each outcome measure were estimated by simulating 50,000 neonates per group.

We first analysed the control group of several RCT designs in order to estimate how commonly used criteria of treatment success would be met. In these cases, we approximated the control arm of a RCT by selecting the AED protocol, outcome measure and criterion for successful seizure reduction that best matched the definitions from the literature: 1) first line AED, success is an 80% seizure reduction from a 1h period before AED compared to a 1h period after AED; 2) second line AED, success is an 80% seizure reduction from 2h before AED compared to 2h period, 2h after AED; 3) second line AED, success is a 50% seizure reduction from a 1h period before AED compared to a 24h period after AED [4, 12, 13]. We simulated the control group of the RCT for a range of T_d (1h to 8h). The proportion of neonates that exceeded the criterion for success was then calculated.

We then analysed the effect of each variable of RCT design independently by simulating subgroups of RCTs where the variable of interest was altered while other variables were fixed. We used a range of T_d from 1h to 8h, three AED protocols, four AED efficacies and five outcome measures resulting in 480 simulated trials.

The sample size with respect to each simulated RCT was considered as a random variable. Changes in sample size due to changes in RCT variable were expressed as proportions or fold changes. All values were summarised using the median, interquartile range and range where applicable. The 95% confidence intervals (CI) of the sample and effect sizes were calculated using bootstrap resampling (1000 iterations for each simulated RCT). In this case, the seizure time course of each neonate in the dataset was considered as one sample. Differences in RCT variables (T_d, outcome measures, AED protocol) were tested using a Wilcoxon signed rank test (a paired test). Finally, we compared the required sample sizes for a subgroup of neonates who had received therapeutic hypothermia to a subgroup who had not received therapeutic hypothermia (normothermic group). This is useful as therapeutic hypothermia has been shown to reduce seizure burden in neonates with HIE and is now the standard of care in many NICUs [19]. Comparisons between therapeutic hypothermia and normothermic groups were performed using Mann Whitney U-tests (an unpaired test).

The trial simulation code was developed with Matlab (Mathworks, Natick, MA, USA) and will be made available on request.

Results

Evidence for the need of a control arm

We found that the placebo control arm of several RCTs had a perceptible rate of apparent success in (Fig 3). This false positive rate was dependent on the definition of the outcome measure, T_d, and the criterion of successful seizure reduction.

Sample Size vs T_d

Tables 1 and 2 show the required sample sizes across RCTs defined by a range of outcome measures, AED protocols and T_d for two target levels of AED efficacy: 80% seizure reduction for 12h and 100% seizure reduction for 72h. Tables 3 and 4 shows the corresponding effect sizes. The median increase in sample size from T_d = 1h to T_d = 8h was 2.1 fold (IQR: 1.7–2.9) across a range of outcome measures, AED protocols and AED efficacies ($p<0.001$; n = 60).

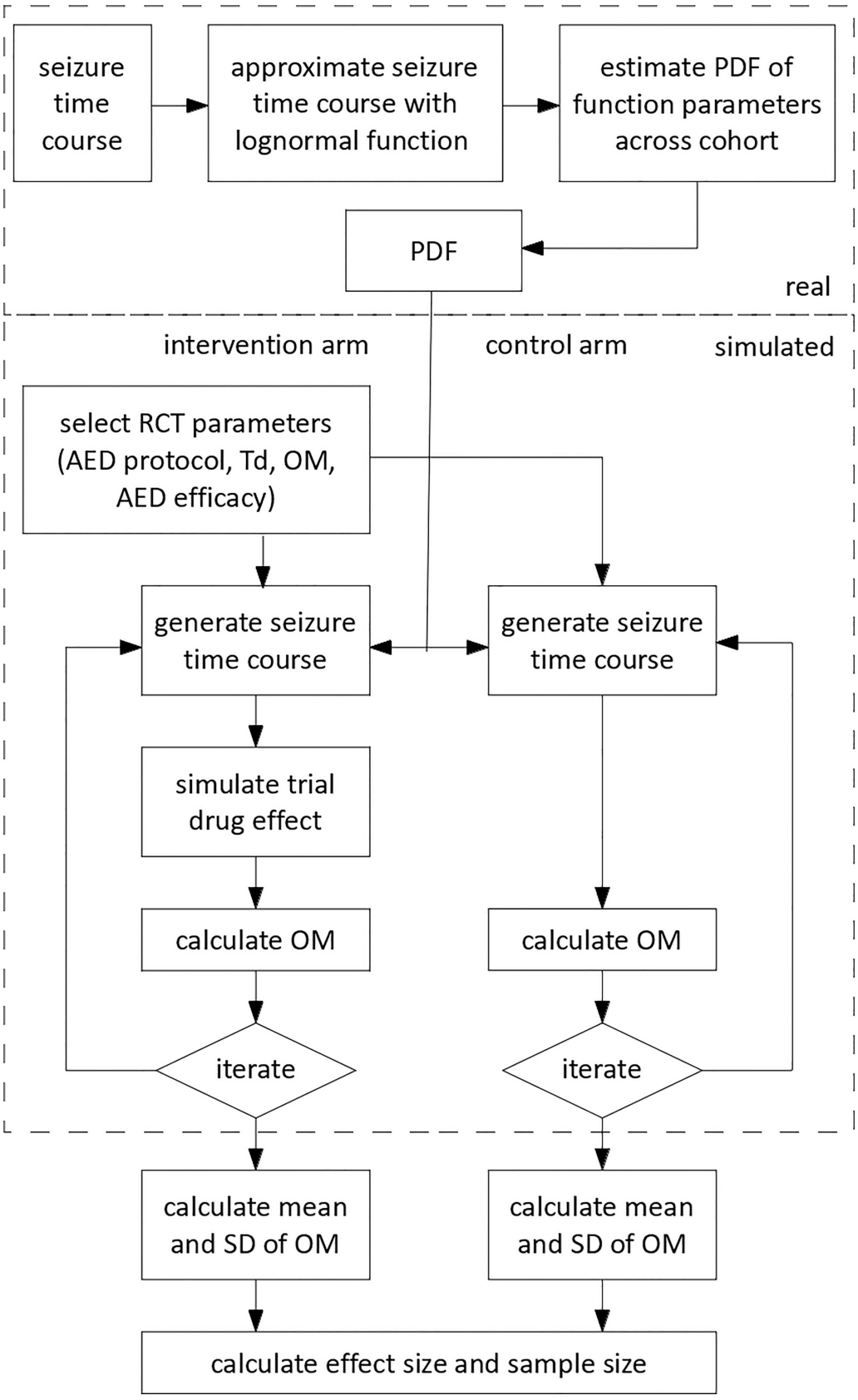

seizure time course
approximate seizure time course with lognormal function
estimate PDF of function parameters across cohort
PDF
real
intervention arm
control arm
simulated
select RCT parameters (AED protocol, Td, OM, AED efficacy)
generate seizure time course
generate seizure time course
simulate trial drug effect
calculate OM
calculate OM
iterate
iterate
calculate mean and SD of OM
calculate mean and SD of OM
calculate effect size and sample size

Fig 2. The RCT simulation used in this study. 1) Real seizure burden (SB) time courses were initially modelled with a lognormal function. The probability density function (PDF) of the lognormal function parameters across a cohort of 41 neonates was estimated. 2) RCT parameters including the AED protocol, delay between seizure onset and intervention (T_d), outcome measure (OM), and level of AED efficacy were selected. A lognormal SB time course for each RCT arm was simulated by selecting lognormal parameters for each arm of the RCT from a multi-variate random variable with the PDF defined previously. The effect of the trial AED (and any other effects according to the AED protocol) was then applied to the SB time course in the intervention arm. The OM was calculated from the SB time course for each RCT arm. This process was iterated across 50000 simulated neonates and the mean and the standard deviation of the OM was then calculated and used to estimate the effect and sample size.

doi:10.1371/journal.pone.0165693.g002

Sample Size vs Outcome Measure

The choice of outcome measure resulted in the largest changes in the required sample size. In general, long term outcome measures resulted in the largest sample sizes and short term measures resulted in the smallest sample sizes. The only exception was for RCTs with a positive control where short term outcome measures had the highest sample size. The choice of outcome measure resulted in a median difference in sample size of, at most, 30.7 fold (IQR: 13.7–40.0) across a range of AED protocols, AED efficacies and delays ($p<0.001$; $n = 96$).

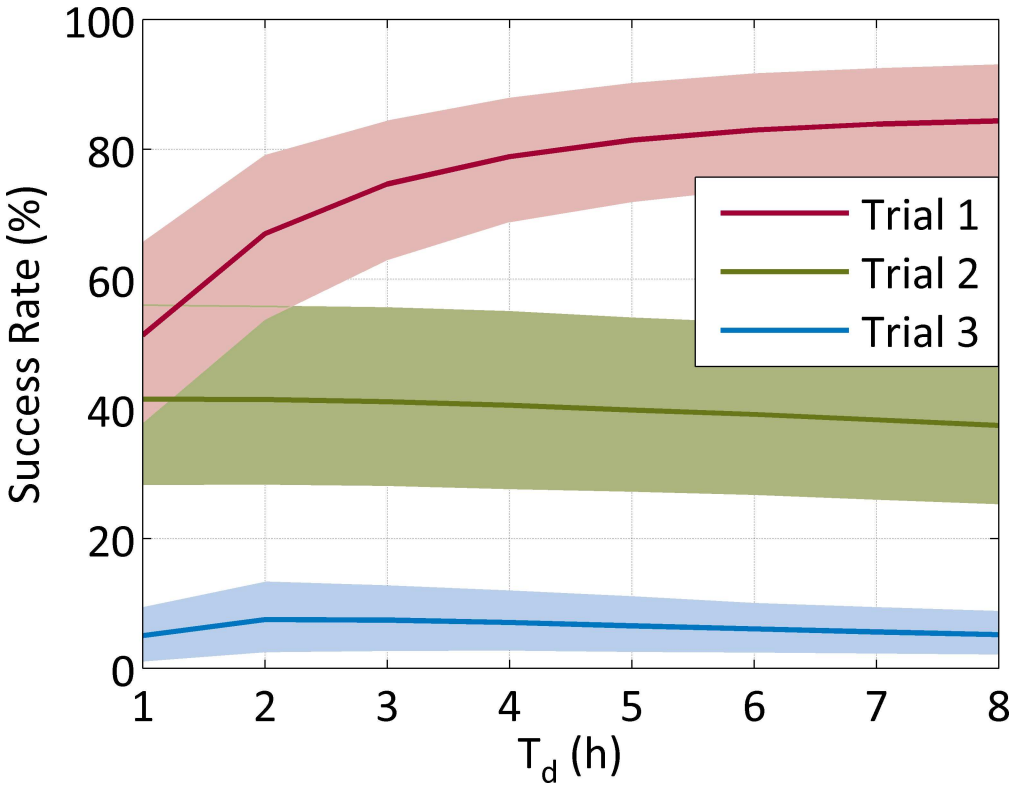


Fig 3. Nominal success rates due to a natural decay in seizure burden in neonates from a placebo control group. Success in trial 1 (second line) 50% reduction in seizure burden measured in a 1h period before placebo compared to the seizure burden in a 24h period after placebo. Success in trial 2 (second line) is an 80% reduction in the seizure burden measured in a 2h period before placebo compared to the seizure burden in a 2h period starting 2h after placebo. Success in trial 3 (first line) is an 80% reduction in the seizure burden measured in a 1h period before placebo compared to the seizure burden in a 1h period after placebo. The shaded areas denote the 95% confidence interval of the estimates.

doi:10.1371/journal.pone.0165693.g003

Table 1. The effect of trial design on sample size with an assumed trial drug efficacy of 80% reduction in seizure burden for 12h. Outcome measures (OM) are total seizure burden (tSB) in minutes, post-intervention seizure burden (pSB) in minutes measured over a duration specified by the subscript in hours, seizure burden response (rSB) in minutes per hour, 1h pre-intervention vs. a post-intervention duration specified by the subscript in hours. Results are presented as sample size (95% CI).

AED protocol	OM	Treatment delay (h)				
		1	2	4	6	8
1st line AED	tSB	268 (136–440)	304 (162–508)	422 (236–724)	606 (348–1056)	870 (518–1508)
placebo control	pSB_1	26(18–34)	32(22–46)	46 (30–70)	58 (36–90)	68 (42–108)
	pSB_{12}	42(26–62)	50(30–76)	64 (38–100)	76 (46–120)	88 (52–140)
	rSB_1	28(18–44)	24(16–36)	32 (20–48)	40 (26–64)	50 (30–80)
	rSB_{12}	98(58–182)	102(62–182)	126 (72–234)	152 (86–288)	170 (100–326)
1st line AED	tSB	532(284–902)	610(328–1058)	844 (474–1476)	1206 (708–2138)	1730(1040–3138)
positive control	pSB_1	608(402–870)	762(538–1190)	1076 (718–1820)	1372 (852–2340)	1654 (992–2860)
	pSB_{12}	64(40–100)	72 (44–116)	90 (54–142)	104 (62–166)	118 (70–192)
	rSB_1	5382 (3028–11158)	7836(4058–17230)	17552 (7004–31878)	23234 (9610–43934)	26444 (10970–67138)
	rSB_{12}	194(110–406)	224 (122–460)	316 (162–652)	390 (208–838)	434 (242–924)
2nd line AED	tSB	446(252–770)	516 (294–886)	716 (428–1230)	1016 (616–1764)	1442 (874–2636)
placebo control	pSB_1	52(32–80)	58 (36–90)	68(42–108)	80 (48–126)	90 (54–144)
	pSB_{12}	70(42–110)	76 (46–120)	88(52–140)	98 (58–160)	108 (64–180)
	rSB_1	36(24–56)	40 (26–64)	50 (30–80)	56 (34–92)	64 (38–106)
	rSB_{12}	140(80–266)	152 (86–288)	170 (100–326)	184 (106–350)	196 (110–372)

doi:10.1371/journal.pone.0165693.t001

Sample size vs AED protocol

In general, RCTs that included comparisons with a control group containing other AEDs (as either positive controls or first line AEDs in a second line trial) required a higher sample size than first line, placebo controlled trials; the median increase in sample size was 3.2 fold (IQR: 1.9–11.9) across a range of AED efficacies, delays and outcome measures ($p<0.001$; n = 320).

Table 2. The effect of trial design on sample size with an assumed trial drug efficacy of 100% reduction in seizure burden for 72h. Outcome measures (OM) are total seizure burden (tSB) in minutes, post-intervention seizure burden (pSB) in minutes measured over a duration specified by the subscript in hours, seizure burden response (rSB) in minutes per hour, 1h pre-intervention vs. a post-intervention duration specified by the subscript in hours. Results are presented as sample size (95% CI).

AED protocol	OM	Treatment Delay (h)				
		1	2	4	6	8
1st line	tSB	40 (26–60)	46 (28–72)	62 (38–100)	86 (52–146)	118 (72–214)
Placebo	pSB_1	16 (10–22)	20 (14–28)	28 (18–44)	36 (22–56)	42 (26–68)
Control	pSB_{12}	26 (16–40)	32 (20–48)	40 (24–62)	46 (28–74)	54 (32–86)
	rSB_1	20 (14–30)	22 (14–30)	30 (20–46)	40 (24–62)	48 (30–76)
	rSB_{12}	62 (36–112)	70 (42–124)	94 (56–176)	118 (68–220)	134 (80–254)
1st line	tSB	48 (30–76)	54 (34–86)	72 (44–116)	98 (60–172)	138 (82–260)
Positive	pSB_1	16 (10–22)	20 (14–28)	28 (18–44)	36 (22–56)	42 (26–68)
Control	pSB_{12}	34 (20–50)	38 (24–58)	44 (28–72)	52 (32–84)	58 (36–96)
	rSB_1	252 (176–380)	396 (256–584)	732 (444–1082)	990 (592–1544)	1198 (710–1906)
	rSB_{12}	102 (58–200)	128 (72–256)	192 (106–388)	244 (136–482)	278 (158–556)
2nd line	tSB	64 (38–102)	72 (44–116)	96 (58–166)	132 (80–244)	182 (106–370)
Placebo	pSB_1	32 (20–50)	36 (22–56)	42 (26–68)	50 (30–78)	56 (34–90)
Control	pSB_{12}	44 (26–68)	46 (28–74)	54 (32–86)	60 (36–100)	68 (40–112)
	rSB_1	34 (22–54)	40 (24–62)	48 (30–76)	54 (34–90)	62 (38–102)
	rSB_{12}	108 (62–200)	118 (68–220)	134 (80–254)	148 (86–278)	158 (90–300)

doi:10.1371/journal.pone.0165693.t002

Table 3. The effect of trial design on effect size with an assumed trial drug efficacy of 80% reduction in seizure burden for 12h. Outcome measures (OM) are total seizure burden (tSB) in minutes, post-intervention seizure burden (pSB) in minutes measured over a duration specified by the subscript in hours, seizure burden response (rSB) in minutes per hour, 1h pre-intervention vs. a post-intervention duration specified by the subscript in hours. The maximum possible effect for each trial is given in the Table 4. The effect size of tSB and pSB_{12} are equal as the assumed efficacy of the trial AED effect is 12h. The effect size of pSB_1 and rSB_1 are also equal as the rSB is measured in minutes per hour and not as a proportion. Results are presented as effect size (95% CI).

AED protocol	OM	Treatment delay (h)				
		1	2	4	6	8
1st line AED	tSB & pSB_{12}	98.1 (54.2–94.9)	91.3 (48.7–91.4)	77.4 (39.5–81.0)	65.1 (31.5–70.4)	54.9 (25.4–60.5)
placebo control	pSB_1 & rSB_1	11.4 (9.2–13.4)	11.3 (8.5–13.6)	10.0 (7.0–12.7)	8.4 (5.6–11.2)	7.0 (4.5–9.7)
	rSB_{12}	8.2 (4.5–7.9)	7.6 (4.1–7.6)	6.3 (3.3–6.7)	5.3 (2.6–5.9)	4.5 (2.1–5.0)
1st line AED	tSB & pSB_{12}	66.9 (29.9–58.4)	61.3 (27.0–54.9)	51.5 (21.5–47.6)	43.5(17.3–40.9)	36.8 (14.0–35.0)
positive control	pSB_1 & rSB_1	0.7 (0.6–0.8)	0.7 (0.5–0.8)	0.6 (0.4–0.8)	0.5 (0.4–0.7)	0.4 (0.3–0.6)
	rSB_{12}	5.6 (2.5–4.9)	5.1 (2.2–4.6)	4.2 (1.8–4.0)	3.5 (1.4–3.4)	3.0 (1.2–2.9)
2nd line AED	tSB & pSB_{12}	70.9 (35.2–75.3)	64.9 (31.5–70.4)	54.8 (25.4–60.5)	46.5 (20.6–51.8)	39.8 (16.9–44.6)
placebo control	pSB_1 & rSB_1	9.2 (6.3–12.0)	8.4 (5.6–11.2)	7.0 (4.5–9.7)	5.9 (3.6–8.4)	4.9 (2.9–7.2)
	rSB_{12}	5.8 (2.9–6.3)	5.3 (2.6–5.9)	4.5 (2.1–5.0)	3.8 (1.7–4.3)	3.3 (1.4–3.7)

doi:10.1371/journal.pone.0165693.t003

Effects on sample size due to therapeutic hypothermia

Required sample sizes were on median 2.6 times greater (IQR: 2.4–3.0) for a RCT based on seizure burden time courses from a subgroup of neonates treated with hypothermia compared to normothermic neonates ($p<0.001$; $n = 480$). More detailed results from the analysis of a subgroup of neonates treated with therapeutic hypothermia is shown in S3 Appendix.

Discussion

Our results suggest that the design of a RCT for assessing the efficacy of an AED in neonates with seizures is challenging with orders of magnitude differences in the required sample size possible depending on the choices of AED protocol, outcome measure, and T_d. The high variability in the magnitude of seizure burden over time within and across neonates translates to

Table 4. The effect of trial design on effect size with an assumed trial drug efficacy of 100% reduction in seizure burden for 72h (the maximum possible effect). Outcome measures (OM) are total seizure burden (tSB) in minutes, post-intervention seizure burden (pSB) in minutes measured over a duration specified by the subscript in hours, seizure burden response (rSB) in minutes per hour, 1h pre-intervention vs. a post-intervention duration specified by the subscript in hours. The effect size of pSB_1 and rSB_1 are equal as the rSB is measured in minutes per hour and not as a proportion. Results are presented as effect size (95% CI).

AED protocol	OM	Treatment Delay (h)				
		1	2	4	6	8
1st line	tSB	212.3 (54.2–94.9)	198.1 (48.7–91.4)	170.8 (39.5–81.0)	147.0 (31.5–70.4)	127.0 (25.4–60.5)
Placebo	pSB_1 & rSB_1	14.2 (9.2–13.4)	14.1 (8.5–13.6)	12.5 (7.0–12.7)	10.5 (5.6–11.2)	8.8 (4.5–9.7)
Control	pSB_{12}	122.7 (54.2–94.9)	114.1 (48.7–91.4)	96.6 (39.5–81.0)	81.1 (31.5–70.4)	68.3 (25.4–60.5)
	rSB_{12}	10.2 (4.5–7.9)	9.5 (4.1–7.6)	8.0 (3.3–6.7)	6.7 (2.6–5.9)	5.7 (2.1–5.0)
1st line	tSB	181.0 (29.9–58.4)	168.1 (27.0–54.9)	145.0 (21.5–47.6)	125.3 (17.3–40.9)	108.9 (14.0–35.0)
Positive	pSB_1 & rSB_1	3.6 (0.6–0.8)	3.5 (0.5–0.8)	3.1 (0.4–0.8)	2.6 (0.4–0.7)	2.2 (0.3–0.6)
Control	pSB_{12}	91.4 (29.9–58.4)	84.2 (27.0–54.9)	70.7 (21.5–47.6)	59.4 (17.3–40.9)	50.2 (14.0–35.0)
	rSB_{12}	7.6 (2.5–4.9)	7.0 (2.2–4.6)	5.8 (1.8–4.0)	4.9 (1.4–3.4)	4.1 (1.2–2.9)
2nd line	tSB	158.3 (35.2–75.3)	146.8 (31.5–70.4)	126.8 (25.4–60.5)	110.1 (20.6–51.8)	96.2 (16.9–44.6)
Placebo	pSB_1 & rSB_1	11.5 (6.3–12.0)	10.5 (5.6–11.2)	8.8 (4.5–9.7)	7.3 (3.6–8.4)	6.1 (2.9–7.2)
Control	pSB_{12}	88.5 (35.2–75.3)	81.1 (31.5–70.4)	68.3 (25.4–60.5)	57.8 (20.6–51.8)	49.3 (16.9–44.6)
	rSB_{12}	7.3 (2.9–6.3)	6.7 (2.6–5.9)	5.7 (2.1–5.0)	4.8 (1.7–4.3)	4.1 (1.4–3.7)

doi:10.1371/journal.pone.0165693.t004

high variability in outcome measures of trial AED efficacy. The present findings offer three practical suggestions for how to optimize a study in terms of minimizing the required sample size and maximizing the validity.

Trials must incorporate a control group. A control group is necessary as simulations show a non-negligible percentage (5–85%) of patients treated with a placebo can fulfil criteria of treatment success due to the natural reduction in seizures over time. This is commonly seen in drug trials for epilepsy [2]. Indeed, it is plausible to speculate that this natural seizure reduction may contribute, in varying degrees, to the reported treatment effects observed in prior uncontrolled studies [13, 21, 22].

Trials should aim to minimise, and control for differences in, T_d between intervention and control groups. A reduction in T_d reduces the required sample size and increases the measured effect. A change in T_d of several hours can double the required sample size. This implies that the entire study protocol needs to proceed very rapidly, advancing from patient identification to recruitment and AED administration within a few hours of seizure onset. This is technically achievable in high level NICUs as modern clinical practice can operate within short time frames as long as early EEG monitoring of all high-risk infants is standard procedure [23, 24]. The dependency of the effect size on T_d indicates that any differences in T_d between intervention and control groups must be controlled for in statistical analyses.

The combination of AED protocol and the choice of outcome measure should be carefully considered as a mismatch may result in an order of magnitude increase in the required sample size. Among outcome measures and AED protocols, post-intervention seizure burden in a first line AED vs. placebo controlled RCT required the lowest required sample size. RCTs using short-term seizure burden response in a positive control trial had the highest required sample size. For other AED protocols, RCTs with the total seizure burden as an outcome measure required large sample sizes. The post-intervention seizure burden was, in general, the outcome measure that resulted in RCTs with the lowest required sample size depending on the post-intervention analysis period. The ability to alter the post-intervention period is advantageous as investigators can nominate a desired trial AED effect time which caters for the incorporation of rescue medications in trial design for seizures that do not respond to treatment or reoccur with significant intensity; medications that if given during the analysis period confound subsequent statistical analyses. Another advantage of the post-intervention seizure burden is that it does not require the monitoring of a baseline period (seizure burden response) or the recording of seizure onset (total seizure burden).

Our present findings are supported by the literature both in terms of studies that did, and did not find, significant differences in seizure burden. In the highly cited and important study of Painter et al. (1999), no significant differences in seizure burden response between phenobarbitone and phenytoin were found in a cohort of 59 neonates [4]. Our findings suggest that this trial was powered to detect differences in seizure efficacy considerably larger than 25% between AEDs (see positive control, first line, rSB_1 in Table 2 – 75% efficacy for positive control vs 100% efficacy for trial drug); in other words, the efficacy of phenobarbitone was not greater than 25% different from phenytoin. A recent uncontrolled study observed AED efficacy with a measure similar to the seizure burden response in a cohort of 19 neonates [17]. In an RCT with a similar design and effect size (placebo control, first line, rSB_1, T_d = 1h in Table 1); our model would have predicted a sample size of 14 neonates per group. Low et al. (2012) also showed a significant reduction in total seizure burden in neonates treated with hypothermia in a cohort of 31 neonates [19]. In this case, our model would predict a cohort of 40 neonates based on a similar effect size; equivalent to an AED effect of 100% for 72h given within 1h of seizure onset (see placebo control, first line, TSB in Table 2). The use of simulation adds further context to

these studies which can be useful for clinicians who are considering altering their treatment protocols in response to RCT findings.

RCTs may be able to reduce the required sample size by assuming more conservative levels of AED efficacy and considering post-hoc stratification of patients according to the intensity of seizure burden. This approach was used successfully, in a recent study that initially found no significant reduction in total seizure burden due to AED treatment managed with EEG monitoring in a cohort of 35 neonates despite a large effect size; however, a significant reduction in total seizure burden was found in a subgroup of neonates with seizures but without status epilepticus [16].

A potential limitation in our work is that only one aetiology (HIE) of neonatal seizures was used as the basis of our simulations. While systematic comparative characterization of seizure time courses is not available across different aetiologies, the present data and clinical experience suggest that seizure time courses may vary with aetiology [9]. This implies that our sample size estimates apply to treatment trials only on neonates with HIE. If sample size estimates for a general cohort of neonates with seizure are required then the use of neonates with HIE provides the best initial data as 1) HIE is the most common aetiology of seizure and 2) the temporal evolution of seizures over time is well defined in this aetiology [6, 7, 10, 25]. The seizures used as a basis for simulations are detected using the visual interpretation of the multi-channel EEG [23]. While this method is the gold standard for seizure detection, it is not perfect and its reliability is reduced when seizures are infrequent or of short duration [26]. We do not expect this to have had a major effect on the results as any variability due to reliability will be vastly outweighed by the variability in seizure time courses across neonates. A technical limitation of our work is the use of simulated seizure burden time courses based on a real but limited dataset of neonates where seizures were treated with AEDs. This reduces the precision of the sample size estimates, however, the relationships between outcome measure, trial protocol, treatment delay, level of trial AED efficacy and sample size are statistically significant.

Our present work shows that it is, in principle, possible to measure the short term efficacy of AEDs in a relatively small cohort of neonates. A more difficult challenge is determining if one AED is more effective than another (RCTs with a positive control); a proposition that more closely achieves clinical equipoise [27]. In order to minimise the sample size, these RCTs require long term outcome measures of seizure burden and must assume a high level of trial AED efficacy far in excess of the current generation of AEDs (see Table 2). This assumption is more critical as modern day NICUs are highly effective positive controls that take advantage of EEG monitoring to target AED therapy and use therapeutic hypothermia for HIE which have, independently, been shown to significantly reduce seizure burden [16, 19].

Supporting Information

S1 Appendix. A summary of the seizure burden in the cohort used to construct a model of seizure time courses.
(DOCX)

S2 Appendix. A model of seizure time courses and trial simulation.
(DOCX)

S3 Appendix. Additional results based on alternate simulations parameters.
(DOCX)

Acknowledgments

NJS was supported by an EU Marie Skłodowska Curie Individual Fellowship (H2020-MCSA-IF-656131), GBB was supported by Science Foundation Ireland (12/RC/2272). SV was

supported by Academy of Finland (#276523 and #288220), Sigrid Juselius Foundation, and the European Union (FP7-HEALTH-2009-4.2–1, grant agreement 241479). The data used in this study was supported by Principal Investigator Awards to GBB from the Health Research Board (RP/2008/238) and the Wellcome Trust (085249). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: NJS GBB SV LHW.

Data curation: NJS.

Formal analysis: NJS.

Funding acquisition: NJS GBB SV LHW.

Investigation: GBB.

Methodology: NJS SV LHW.

Project administration: GBB SV.

Resources: GBB SV NJS.

Software: NJS.

Supervision: SV.

Validation: NJS.

Visualization: NJS SV LHW.

Writing – original draft: NJS.

Writing – review & editing: NJS GBB SV LHW.

References

1. Vandevanter DR, Yegin A, Morgan WJ, Millar SJ, Pasta DJ, Konstan MW. Design and powering of cystic fibrosis clinical trials using pulmonary exacerbation as an efficacy endpoint. J Cyst Fibros 2011; 10: 453–459. doi: 10.1016/j.jcf.2011.07.003 PMID: 21803665
2. Goldenholz DM, Moss R, Scott J, Auh S, Theodore WH. Confusing placebo effect with natural history in epilepsy: A big data approach. Ann Neurol 2015; 78: 329–336. doi: 10.1002/ana.24470 PMID: 26150090
3. Cock HR. Established status epilepticus treatment trial (ESETT). Epilepsia 2011; 52 S8:50–52.
4. Painter MJ, Scher MS, Stein AD, Armatti S, Wang Z, Gardiner JC et al. Phenobarbital compared with phenytoin for the treatment of neonatal seizures. N Engl J Med 1999; 341:485–489. doi: 10.1056/NEJM199908123410704 PMID: 10441604
5. Boylan GB, Rennie JM, Pressler RM, Wilson G, Morton M, Binnie CD. Phenobarbitone, neonatal seizures, and video-EEG. Arch Dis Child-Fetal 2002; 86: F165–F170.
6. Wusthoff CJ, Dlugos DJ, Gutierrez-Colina A, Wang A, Cook N, Donnelly M et al. Electrographic seizures during therapeutic hypothermia for neonatal hypoxic-ischemic encephalopathy. J Child Neurol 2011; 26: 724–728. doi: 10.1177/0883073810390036 PMID: 21447810
7. Lynch NE, Stevenson NJ, Livingstone V, Murphy BP, Rennie JM, Boylan GB. The temporal evolution of electrographic seizure burden in neonatal hypoxic ischemic encephalopathy Epilepsia 2012: 53; 549–557. doi: 10.1111/j.1528-1167.2011.03401.x PMID: 22309206
8. Sánchez SM, Arndt DH, Carpenter JL, Chapman KE, Cornett KM, Dlugos DJ et al. Electroencephalography monitoring in critically ill children: Current practice and implications for future study design. Epilepsia 2013; 54: 1419–1427. doi: 10.1111/epi.12261 PMID: 23848569

9. Low E, Mathieson SR, Stevenson NJ, Livingstone V, Ryan CA, Bogue CO et al. Early postnatal EEG features of perinatal arterial ischaemic stroke with seizures. PLoS One 2014: 7 pages, e100973 doi: 10.1371/journal.pone.0100973 PMID: 25051161
10. Lynch NE, Stevenson NJ, Livingstone V, Mathieson S, Murphy BP, Rennie JM et al. The temporal characteristics of seizures in neonatal hypoxic ischemic encephalopathy treated with hypothermia, Seizure 2015; 33: 60–65. doi: 10.1016/j.seizure.2015.10.007 PMID: 26571073
11. Clancy RR. Summary proceedings from the neurology group on neonatal seizures. Pediatrics 2006; 117 (S1): S23–S27.
12. Pressler RM, Boylan GB, Marlow N, Blennow M, Chiron C, Cross JH et al. Bumetanide for the treatment of seizures in newborn babies with hypoxic ischaemic encephalopathy (NEMO): an open-label, dose finding, and feasibility phase 1/2 trial. Lancet Neurol 2015; 14, 469–477. doi: 10.1016/S1474-4422(14)70303-5 PMID: 25765333
13. Abend NS, Gutierrez-Colina AM, Monk HM, Dlugos DJ, Clancy RR. Levetiracetam for treatment of neonatal seizures. J Child Neurol 2011; 26: 465–470. doi: 10.1177/0883073810384263 PMID: 21233461
14. Björkman ST, Miller SM, Rose SE, Burke C, Colditz PB. Seizures are associated with brain injury severity in a neonatal model of hypoxia–ischemia. Neuroscience 2010; 166: 157–167. doi: 10.1016/j.neuroscience.2009.11.067 PMID: 20006975
15. van Rooij LGM, Toet MC, van Huffelen AC, Groenendaal F, Laan W, Zecic A et al. Effect of treatment of subclinical neonatal seizures detected with aEEG: randomized, controlled trial. Pediatrics 2010; 125: e358–e366. doi: 10.1542/peds.2009-0136 PMID: 20100767
16. Srinivasakumar P, Zempel J, Trivedi S, Wallendorf M, Rao R, Smith B et al. Treating EEG Seizures in Hypoxic Ischemic Encephalopathy: A Randomized Controlled Trial. Pediatrics 2015; 136: e1302–e1309. doi: 10.1542/peds.2014-3777 PMID: 26482675
17. Low E, Stevenson NJ, Mathieson SR, Livingstone V, Ryan CA, Rennie JM et al. Short-term effects of phenobarbitone on electrographic seizures in neonates. Neonatalogy 2016; 110:40–46
18. Murray DM, Boylan GB, Ali I, Ryan CA, Murphy BP, Connolly S. Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures. Arch Dis Child-Fetal 2008; 93: F187–F191.
19. Low E, Boylan GB, Mathieson SR, Murray DM, Korotchikova I, Stevenson NJ et al. Cooling and seizure burden in term neonates: an observational study. Arch Dis Child-Fetal 2012; 97: F267–F272.
20. Desu MM, Raghavarao D. Sample Size Methodology. San Diego, USA: Academic Press; 1990, pp. 30.
21. Hellström-Westas L, Westgren U, Rosén I, Svenningsen NW. Lidocaine for treatment of severe seizures in newborn infants. Acta Paediatrica, 1988; 77: 79–84.
22. van den Broek MPH, Rademaker CMA, van Straaten HLM, Huitema AD, Toet MC, de Vries LS et al. Anticonvulsant treatment of asphyxiated newborns under hypothermia with lidocaine: efficacy, safety and dosing. Arch Dis Child-Fetal 2013; 98: F341–F345.
23. Boylan GB, Stevenson NJ, Vanhatalo S. Monitoring neonatal seizures. Semin Fetal Neonatal Med 2013; 18: 202–208. doi: 10.1016/j.siny.2013.04.004 PMID: 23707519
24. Glass HC. Neonatal seizures: advances in mechanisms and management. Clin Perinatol 2014; 41: 177–190. doi: 10.1016/j.clp.2013.10.004 PMID: 24524454
25. Vasudevan C,Levene M. Epidemiology and aetiology of neonatal seizures. Semin Fetal Neonatal Med 2013; 18: 185–191. doi: 10.1016/j.siny.2013.05.008 PMID: 23746578
26. Stevenson NJ, Clancy RR, Vanhatalo S, Rosén I, Rennie JM, Boylan GB. Interobserver agreement for neonatal seizure detection using multichannel EEG, Ann Clin Transl Neurol 2015; 2: 1002–1011. doi: 10.1002/acn3.249 PMID: 26734654
27. Freedman B. Equipoise and the ethics of clinical research. New Eng J Med 1987; 317: 141–145. doi: 10.1056/NEJM198707163170304 PMID: 3600702