Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

Interim sample size reestimation for adequately powered series of N-of-1 trials

Daphne N. Weemering*

[1]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

Correspondence
*Corresponding author name, This is sample corresponding address. Email: authorone@gmail.com

Present Address
Padualaan 14, 3584 CH Utrecht, The Netherlands

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean ut elit odio. Donec fermentum tellus neque, vitae fringilla orci pretium vitae. Fusce maximus finibus facilisis. Donec ut ullamcorper turpis. Donec ut porta ipsum. Nullam cursus mauris a sapien ornare pulvinar. Aenean malesuada molestie erat quis mattis. Praesent scelerisque posuere faucibus. Praesent nunc nulla, ullamcorper ut ullamcorper sed, molestie ut est. Donec consequat libero nisi, non semper velit vulputate et. Quisque eleifend tincidunt ligula, bibendum finibus massa cursus eget. Curabitur aliquet vehicula quam non pulvinar. Aliquam facilisis tortor nec purus finibus, sit amet elementum eros sodales. Ut porta porttitor vestibulum. Integer molestie, leo ut maximus aliquam, velit dui iaculis nibh, eget hendrerit purus risus sit amet dolor. Sed sed tincidunt ex. Curabitur imperdiet egestas tellus in iaculis. Maecenas ante neque, pretium vel nisl at, lobortis lacinia neque. In gravida elit vel volutpat imperdiet. Sed ut nulla arcu. Proin blandit interdum ex sit amet laoreet. Phasellus efficitur, sem hendrerit mattis dapibus, nunc tellus ornare nisi, nec eleifend enim nibh ac ipsum. Aenean tincidunt nisl sit amet facilisis faucibus. Donec odio erat, bibendum eu imperdiet sed, gravida luctus turpis.

KEYWORDS:
N-of-1 trials; Sample size reestimation; Simulation study;

1 | INTRODUCTION

Randomized controlled trials (RCTs) are considered the gold standard in determining treatment efficacy in healthcare. At first glance, these standard RCTs seem to earn their position as the randomization of patients into a parallel experimental and control condition works quite well in controlling factors that are not under experimental control, creating the possibility to make causal inferences about the usefulness of the intervention. A drawback, however, is that these standard RCTs require a relatively large sample size to establish the effectiveness of treatment with sufficient power. In all instances it is desirable to limit the number of subjects to enter a medical study, as it exposes individuals to potentially inferior treatment. However, for the instances of finding the right intervention for patients with rare diseases, standard RCTs become unfeasible due to the relatively large sample size they require. As populations of patients with rare diseases only have a limited number of people that could enter such a experimental study, alternatives should be sought to obtain reliable estimates of the treatment effect.

A clinical trial methodology that offers the possibility to reduce the number of subjects necessary to find a treatment effect with sufficient power, is the N-of-1 trial. The N-of-1 trial is a randomized controlled multiple crossover trial where a single patient repeatedly receives the experimental and control intervention in a random order, spread out over multiple cycles[1]. As

the experiment is conducted within a single patient, the advantage of the N-of-1 trial is that a patient-specific treatment effect estimate is obtained. Often, clinical evidence that is generated by standard RCTs turns out to have poor generalization and is therefore to a limited amount applicable to patients in ordinary practices[2]. Also, treatments that are shown to be safe on average, may have a disbalance in risks and benefits to individual patients[3]. With an N-of-1 trial, these issues can be avoided by purely estimating the effect of treatment on a single patient.

To obtain reliable results, N-of-1 trials should be performed under specific clinical circumstances. First of all, the disease under study should be long term and stable over time in order to avoid that the treatment differences will be obscured within and between cycles due to disease progression. Second, the intervention should not modify the course of the disease and should have a rapid on- and offset of biological action of the medication[4,5].

A single N-of-1 trial does not compare itself with a standard RCT, as the results of a single trial are specific to an individual patient, and hence, cannot be generalized to the population. However, when the results of several separate N-of-1 trials are combined it is possible to estimate the population treatment effect[6]. Even though every patient is unique, there is also some similarity in the disease progression and the treatment response of patients. In the combined analysis of separate N-of-1 trials, both the magnitude of the treatment effect as well as the heterogeneity in treatment response are taken into account[6]. The combined estimate of the separate trials forms the basis of knowledge about the treatment, whereas the estimate of the heterogeneity in treatment response allows for determining optimal treatment for every single patient by balancing the similarities and differences. As it is possible to estimate the population treatment effect within each subject, combined N-of-1 trials can serve as efficient alternatives for standard parallel RCTs.

Combined N-of-1 trials, now referred to as series of N-of-1 trials, have been performed in, inter alia, studying the influence of mexiletine on nondystrophic myotonia[7], studying the effectiveness of methylphenidate on fatigue in patients with end-stage cancer[8], an for investigating the usefulness of sildenafil on Raynaus-Phenomenon patients[9]. Reasons for choosing the N-of-1 trial methodology vary, in general but also specifically for these aforementioned studies. The latter study chose to conduct a series of N-of-1 trials due to the heterogeneity in treatment response that should be taken into account, whereas the first two studies chose for the N-of-1 trial methodology due to inability to achieve the required sample size for a standard RCT.

Series of N-of-1 trials require a smaller sample size compared to standard two-arm RCTs, because every patient serves as its own control, creating the possibility to obtain multiple observations per individual[4]. A prior sample size calculations are necessary to avoid under- or overpowering the study, for planning on allocating resources and for assessing the feasibility of the study. For these a priori sample size calculations, assumptions have to made with regard to unknown parameters in the model. The nuisance parameters (i.e., the stochastic model components) are often unknown before the start of the studies. Incorrect estimates for these nuisance parameters can lead to substantial over- or underpowering. Overpowering a study potentially exposes too many patients to inferior treatment, whereas underpowering a study increases the risk of committing an error of the second kind, where one is not able to find an effect of treatment when there exists one in the population.

Sample size formulas have been derived for series of N-of-1 trials for both random and fixed effects models[10]. As the main objective of combining the results of separate N-of-1 trials is to make inferences to the population, random effects models are most appropriate and of interest here. In series of N-of-1 trials, the within- and between subject variance of the response to treatment are unknown at the start of the studies[10]. Sample size estimation in series of N-of-1 trials require estimates for these unknown parameters. Taking estimates of nuisance parameters from other studies can be unreliable because of differences in the study population, background conditions or study design[11]. Furthermore, an estimate of the between-subject variance in treatment response is often not available because the kind of study to obtain these estimates is a trial incorporating such a component, such as a series of N-of-1 trials[12]. Series of N-of-1 trials are not (yet) that common, and even if similar series of N-of-1 trials exist, these kinds of estimates are usually not reported in the literature.

An appealing strategy for conquering the problem of incorrect assumptions for unknown parameters in sample size calculations is the two-stage design. With this design, the required sample size is calculated by making reasonable assumptions for the unknown parameters. Then, data is obtained up until a predetermined interim point along the trial. The data that is obtained up until the interim point is used to estimate the parameters that were unknown prior to the studies. These estimates at interim are then used to reestimate the sample size, which can then be adjusted accordingly[13]. This two-stage design, now referred to as interim sample size reestimation, can avoid a study to become under- or overpowered in the scenario where wrong assumptions are made with regard to unknown parameters in sample size calculations. A distinction can be made between interim sample size reestimation based on nuisance parameter estimates and based on treatment effect estimates[14]. This thesis will not cover the latter approach, but rather focuses on interim sample size reestimation based on nuisance parameter estimates.

A general concern with interim sample size reestimation, is the inflation of the type I error rate. Various studies have attempted to analytically calculate or control the type I error rate under various circumstances [15,16,17,18]. Under specific circumstances, this type I error rate becomes inflated, and this is not desirable. Investigating the influence of interim sample size reestimation for series of N-of-1 trials on the type I error rate is not the main objective of this thesis, but it will be assessed whether the type I error rate becomes inflated for specific scenarios considered here.

The application of interim sample size reestimation has not yet been investigated in the context of series of N-of-1 trials and no specific guidelines have been established. Moreover, the minimally required sample size for interim sample size reestimation in series of N-of-1 trials has also not yet been established. With the use of simulation studies, this thesis aims at establishing guidelines for the minimally required sample size for sample size reestimation in series of N-of-1 trials, and to compare series of N-of-trials incorporating interim sample size reestimation with series of N-of-1 trials having a fixed sample size. Power and expected sample size are important measures of performance herein. Furthermore, the type I error rate between series of N-of-1 trials with and without the incorporation of interim sample size reestimation will be assessed.

The remainder of this paper will be structured as follows: In sections 2, the model used for the simulation studies and the notation are discussed. In section 3, the setup of the simulation studies will be explained. Section 4 discusses the results of the simulations studies, and section 5 includes the conclusions that can be drawn based on the results, the limitations of this study and considerations for future research on this topic.

References

1. Guyatt G, Sackett D, Taylor DW, Ghong J, Roberts R, Pugsley S. Determining optimal therapy—randomized trials in individual patients. *New England Journal of Medicine* 1986; 314(14): 889–892.
2. Greenfield S, Kravitz R, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *The American journal of medicine* 2007; 120(4): S3–S9.
3. Duan N, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of clinical epidemiology* 2013; 66(8): S21–S28.
4. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. *PloS one* 2016; 11(12): e0167167.
5. Nikles J, Mitchell GK, Schluter P, et al. Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. *Journal of clinical epidemiology* 2011; 64(5): 471–480.
6. Zucker D, Schmid C, McIntosh M, D'agostino R, Selker H, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of clinical epidemiology* 1997; 50(4): 401–410.
7. Stunnenberg BC, Raaphorst J, Groenewoud HM, et al. Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated N-of-1 trials. *Jama* 2018; 320(22): 2344–2353.
8. Mitchell GK, Hardy JR, Nikles CJ, et al. The effect of methylphenidate on fatigue in advanced cancer: an aggregated N-of-1 trial. *Journal of pain and symptom management* 2015; 50(3): 289–296.
9. Roustit M, Giai J, Gaget O, et al. On-demand sildenafil as a treatment for Raynaud phenomenon: a series of N-of-1 trials. *Annals of Internal Medicine* 2018; 169(10): 694–703.
10. Senn S. Sample size considerations for n-of-1 trials. *Statistical methods in medical research* 2019; 28(2): 372–383.
11. Zucker DM, Denne J. Sample–size redetermination for repeated measures studies. *Biometrics* 2002; 58(3): 548–559.
12. Senn S. Mastering variation: variance components and personalised medicine. *Statistics in medicine* 2016; 35(7): 966–977.
13. Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. *Journal of biopharmaceutical statistics* 2005; 15(4): 559–574.

14. Proschan MA. Sample size re-estimation in clinical trials. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 2009; 51(2): 348–357.

15. Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics* 1945; 16(3): 243–258.

16. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in medicine* 1990; 9(1-2): 65–72.

17. Kieser M, Friede T. Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in medicine* 2000; 19(7): 901–911.

18. Gao P, Ware JH, Mehta C. Sample size re-estimation for adaptive sequential design in clinical trials. *Journal of Biopharmaceutical Statistics* 2008; 18(6): 1184–1196.