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; statistical methods;

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 balancing factors that are not under experimental control, allowing for unbiased estimation of the population treatment effect. A drawback, however, is that these standard RCTs require a relatively large sample size to establish the effectiveness of treatment with sufficient power. For the instances of finding the right intervention for patients with rare diseases, i.e., small patient populations, standard RCTs become therefore unfeasible.

The N-of-1 trial design can offer a solution for clinical research in small patient populations. A N-of-1 trial is a randomized controlled multiple crossover trial where a single patient repeatedly receives the experimental and control intervention in multiple cycles, where the allocation of the interventions in each cycle is in a random order[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. This allows these single patient trials to identify the best treatment for each patient[2].

There are some essential clinical conditions that are suitable for the use of a N-of-1 trial. First, the medical condition for which the intervention in prescribed should be chronic and (relatively) stable over time in order to reduce the chance that the progression of the disease can obscure the treatment differences between and within the trial cycles[3,4]. Moreover, the intervention being tested in the N-of-1 trial should have a rapid on- and offset of biological action, and should have a short half-life to ensure that there is rapid washout as the cycles alternate[4]. Additionally, the effect of the intervention should be measured using a validated (clinical) outcome measure (e.g., choosing the right scale ensuring that real benefits and real burdens are being measured). Lastly, the intervention used in the study should not alter the underlying condition, as this will make it unable to interpret the results for an individual as the trial progresses[4]. This all necessitates careful selection of participants, short time cycles and relatively stable symptoms.

As results of a single N-of-1 trial are specific to an individual patient and can therefore not be generalized to the population, a single N-of-1 trial does not compare itself with a standard RCT. However, combining several individual N-of-1 trials, under the condition that the trials are identical, creates the possibility to estimate the population-level treatment effect[5]. In the combined analysis of separate N-of-1 trials, now referred to a series of N-of-1 trials, both the magnitude of the average treatment effect as well as the heterogeneity in treatment response are taken into account[5]. Comparing multiple treatment cycles combined with the recognition that the variability in response within individuals is typically lower than the variability between individuals, a smaller sample size is required to detect an effect of treatment in series of N-of-1 trials compared to parallel RCTs[4]. This makes series of N-of-1 trials a valuable alternative to standard RCTs in the accumulation of a comprehensive evidence base in populations of people with rare diseases.

These series of N-of-1 trials have been performed for, among others, studying the effect of mexiletine on nondystrophic myotonia[6], studying the effectiveness of methylphenidate on fatigue in patients with end-stage cancer[7], and for investigating the usefulness of sildenafil on Raynaus-Phenomenon patients[8]. 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.

As in any clinical study, a priori sample size determination is necessary to avoid under- or overpowering the study, for planning on allocating resources and for assessing the feasibility of the study. Sample size formulas have been derived for series of N-of-1 trials for both random and fixed effects models[9]. As the main objective of combining the results of separate N-of-1 trials is to make inferences regarding the population treatment effect, random effects models are most appropriate and of interest here. For the sample size calculations, assumptions have to be made with regard to the parameters in the model, such as the clinically relevant difference and the nuisance parameters. However, the nuisance parameters in the model, which concern the within- and between subject variance of the response to treatment for series of N-of-1 trials[10], are generally unknown at the start of the study. 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 (or trials) 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. Making unrealistic assumptions for these nuisance parameters can lead to substantial over- or underpowering, where the former exposes too many patients to potentially inferior treatment and the latter increases the risk of failing to identify a clinically relevant treatment effect due to a lack of power.

An appealing strategy for conquering the problem of incorrect assumptions for unknown parameters in sample size calculations is a two-stage design with interim sample size reestimation based on nuisance parameter estimates. With this design, the initially required sample size is calculated by making reasonable assumptions for the unknown nuisance parameters. Then, a portion of the data is collected up until a prespecified interim point along the trial and the unknown nuisance parameters are estimated using the data observed so far. These estimates are then used to update the power analysis and to adjust the sample size. Subsequently, the study is continued until the adjusted sample size is reached, and finally, the hypothesis is tested with all the data[13]. Simulations studies have shown that this method has a high potential to protect from an incorrect sample size if the nuisance parameters where misspecified at the design stage of the study for standard RCTs[14]. A distinction can be made between interim sample size reestimation based on nuisance parameter estimates and based on treatment effect estimates[15]. This thesis will cover the first approach.

A concern with interim sample size reestimation based on estimates of nuisance parameters, is the inflation of the type I error rate[16,14,17]. 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. However, it will be assessed whether the type I error rate becomes inflated for specific scenarios considered here, allowing future research to build upon these results.

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. With this thesis, the usefulness and feasibility of interim sample size reestimation in series of N-of-1 trials will be investigated. With the use of simulation studies, interim sample size reestimation in series of N-of-1 trials will be compared with a similar design having a fixed sample size. Furthermore, the type I error rate for series of N-of-1 trials with interim sample size reestimation will be evaluated. Finally, the minimally required sample size that is necessary for reliable interim sample size reestimation will be determined.

The remainder of this thesis will be structured as follows: In section 2, notation, sample size calculations for series of N-of-1 trials, and the model that is used for the simulation studies are discussed. In section 3 the design of the simulation studies in which power, type I error rate and the expected sample size are evaluated will be explained. Section 4 discusses the results of the simulations studies. And finally, this thesis is concluded with a discussion which is outlined in section 5.

2 | METHODOLOGY

First, the methodology of a one-stage design for series of N-of-1 trials will be discussed in section 2.1, in which notation, the model used and sample size calculations for series of N-of-1 trials are introduced. In section 2.2, the procedure for interim sample size reestimation in series of N-of-1 trials will be outlined.

2.1 | One-stage design

In a series of N-of-1 trials, n subjects in k cycles, with each cycle consisting of two periods, receive the experimental condition in one period and the control condition in the other period. The order of treatment administration within each cycle will be randomly determined. At the end of each period, the outcome Y_{ijt} is measured, indicating the outcome for patient i ($i = 1, ..., n$) in cycle j ($j = 1, ..., k$) who is given treatment t ($t = 1, 2$). It is assumed that the disease under study is stable over time, that carryover effects are absent because of a sufficient duration of the washout period, and that there are no missing data. Furthermore, it is assumed that the outcome is a continuous measure and that it is normally distributed according to the following model:

$$Y_{ijt} = \lambda_i + \beta_{ij} + \epsilon_{ijt} + Z_{ijt}\tau_i \tag{1}$$

In this model[10], $\lambda_i \sim N(\Lambda, \phi^2)$, $\beta_{ij} \sim N(0, \gamma^2)$, $\epsilon_{ijt} \sim N(0, \sigma^2)$ and $\tau_i \sim N(T, \psi^2)$. $Z_{ijt} = \frac{1}{2}$ or $-\frac{1}{2}$, dependent on whether $t = 1$ or 2 for patient i in cycle j. In this model, λ_i represents the random effect for subject i, β_{ij} indicates the cycle effects for subject i in cycle j, ϵ represents the i-th subject's random error for the j-th cycle and treatment t, and finally, τ_i indicates the treatment effect for subject i. In this model, all the nuisance parameters are assumed to be independent of each other.

However, under the assumption that the data is balanced, a simpler model for the treatment differences for subject i in cycle j can be derived from equation 1 by subtracting the values from the first period in every cycle from the second period in the cycle, and subsequently divide by $(Z_{ij1} - Z_{ij2})$[10]:

$$d_{ij} = \tau_i + \epsilon_{ij} \tag{2}$$

Here, d_{ij} represents the observed treatment difference for patient i in cycle j, where treatment 1 is consistently subtracted from treatment 2. In this model, $\tau_i \sim N(T, \psi^2)$, as before and $\epsilon_{ij} \sim N(0, 2\sigma^2)$. τ_i is the random treatment effect for patient i, and ϵ_{ij} are random within-subject within-cycle disturbance terms. These disturbance terms are assumed to be i.i.d. (independent and identically distributed) both across cycles and across patients. Furthermore, τ_i and ϵ_{ij} are assumed to be independent of each other. The variance term of ϵ_{ij}, $2\sigma^2$, is given to make this model compatible with a mixed model using the original observations[9].

2.1.1 | Sample size calculations in N-of-1 trials

From the model of the treatment differences in equation 2, the average treatment effect in the population and the corresponding variance can be derived, both necessary for calculating the required sample size. Following Senn[9], an average treatment effect for each patient can be obtained:

$$\bar{d}_{i.} = \frac{\sum_{j=1}^{k} d_{ij}}{k} \tag{3}$$

The average over all the n averages is equal to $\hat{T} = \sum_{i=1}^{n} \sum_{j=1}^{k} d_{ij}/nk$, which can be used to test the differences between the two treatments under investigation. This estimate has a variance of $var(\hat{T}) = \psi^2 + 2\sigma^2/kn$. The variance at the patient level, the variance of the n estimates of $\bar{d}_{i.}$, is then defined as:

$$var(\bar{d}_{i.}) = \psi^2 + 2\sigma^2/k \tag{4}$$

This estimate has $(n-1)$ degrees of freedom[9] and can be used to calculate the required sample size to achieve the desired power under hypothesized values for the nuisance parameters, ψ^2 and σ^2, now denoted as ψ_h^2 and σ_h^2, the clinically relevant difference Δ, and the two-sided significance level α using a one sample t-test. For the calculation of the sample size, the standard deviation of the estimate in equation 4 is used. The exact formula for the calculation of the sample size is based on a non-central t-distribution. Computation of the sample size therefore requires an iterative process and this can straightforwardly be done with the `pwr.t.test` function of the `pwr` package[18] in `R`[19].

2.2 | Two-stage design

To cope with the problem around the a priori uncertainty regarding the nuisance parameters in the model, interim sample size reestimation is applied. For this process, the following steps are taken[14,16,20]:

1. Specify the clinically relevant difference (Δ), the type I error rate (α), the desired power ($1-\beta$), the proportion of the initial sample size on which interim sample size reestimation will be based (f) and the a priori hypothesized estimates for the nuisance parameters (ψ_h^2 and σ_h^2).
2. Use ψ_h^2 and σ_h^2 to estimate the initial sample size n_{init} that yields the desired level of power. All sample sizes are rounded up.
3. Use $f n_{init} = n_{frac}$, the initial sample size on which interim sample size reestimation is based, to estimate $\hat{\sigma}^2$ and $\hat{\psi}^2$.
4. Then, use $\hat{\sigma}^2$ and $\hat{\psi}^2$ to find the new total sample size, n_{final}, that is needed to achieve the target power and subsequently observe the additional $n_{final} - n_{frac}$ patients. If $n_{final} - n_{frac} \leq 0$, n_{frac} is used as the final sample size. Also, if the effect size (Cohen's d, mean treatment difference divided by the standard deviation) for recalculating the sample size at interim becomes larger than 10 due to small $\hat{\sigma}^2$ and $\hat{\psi}^2$, the sample size becomes too small. In that case, the effect size will be set equal to 10.
5. Test the hypothesis on all the $n_{frac} + (n_{final} - n_{frac})$ observations.

3 | SIMULATIONS

Two simulation studies are performed to (I) compare the power and average sample size of trials incorporating interim sample size reestimation with series of N-of-1 trials having a fixed sample size, and (II) to evaluate the type I error rate for series of N-of-1 trials that incorporate interim sample size reestimation. When the effect of interim sample size reestimation in series of N-of-1 trials on power and the type I error rate is established, the minimally required sample size that is necessary for reliable interim sample size reestimation will be determined.

The linear mixed model provided in equation 2 will be used to simulate data for a series of N-of-1 trials. For the simulation studies, various scenarios are considered, each under different combinations of ψ_h^2, σ_h^2, f, and for the actual values of the nuisance parameters which are used for generating the data, ψ_t^2 and σ_t^2. This eventually results in $3^5 = 243$ different scenarios in each simulation study. All the relevant parameter values considered in the simulation studies are displayed in table 1.

In the first simulation study, power will be evaluated and compared between series of N-of-1 trials incorporating interim sample size reestimation and series of N-of-1 trials that have a fixed sample size. Power is computed as the proportion of

TABLE 1 Parameters in the two simulation studies.

Description	Values
Design parameters	
Fraction of initial sample size	$f = 0.25, 0.5, 0.75$
Cycles per patient	$k = 3$
Power	$1 - \beta = 0.8$
Two-sided nominal significance level	$\alpha = 0.05$
Clinically relevant difference	$\Delta = 0, 1$
Model parameters	
Within-patient within-cycle variance (data generation)	$\sigma_t^2 = 0.25, 0.5, 1$
Variance of treatment effect (data generation)	$\psi_t^2 = 0.5, 1, 2$
Within-patient within-cycle variance (hypothesized)	$\sigma_h^2 = 0.25, 0.5, 1$
Variance of treatment effect (hypothesized)	$\psi_h^2 = 0.5, 1, 2$
Average treatment effect	$T = \Delta = 0, 1$
Simulation parameter	
Number of simulations	$N = 10000$
Outcome parameters	
Statistical power	Proportion of iterations that the null hypothesis is rejected under $T = 1$ (for data generation)
Type I error rate	Proportion of iterations that the null hypothesis is rejected under $T = 0$ (for data generation)
Average sample size	Average sample size under reestimation for the N iterations

iterations that the null hypothesis of no treatment effect is rejected ($p \leq \alpha$) when $T = 1$, both for data generation and for calculating the initial sample size. In the second simulation study, the type I error rate, α, will be evaluated for series of N-of-1 trials. As α is the probability of falsely rejecting the null hypothesis when it is true[21], setting $T = 0$ for data generation and applying the same calculation as for the power, the estimated α-level should approximate the nominal rate. Both simulation studies compute the average reestimated sample size, which provides an opportunity to see if there should be certain requirements for the minimum sample size on which re-estimation is based. To do so, the average reestimated sample sizes will be compared with the corresponding true sample size. Power and the type I error rate are also involved in deciding on these requirements; they should at least meet the prespecified 0.8 and 0.05, respectively.

Linear mixed models will be fitted to the simulated data using the `lme4` package[22] in `R` (version 4.1.3)[19]. Because treatment effects are modeled within cycles, an intercept-only model will suffice to estimate the mean treatment effect and patient-specific random effects. Restricted maximum likelihood (REML) will be applied to estimate the variance parameters, as this approach, to the contrary of "regular" maximum likelihood (ML), produces unbiased estimates for the variance components in the model for smaller sample sizes[23]. Sample size calculations for n_{init} and n_{final} will be performed using a two tailed one sample t-test from the `pwr.t.test` function from the `pwr` package in `R`[18]. If the effect size (Cohen's d, the standardized effect size) for (re)calculating the sample size becomes larger than 11 due to small $\hat{\sigma}^2$ and $\hat{\psi}^2$, the sample size becomes too small (approximately 2). In that case, the effect size will be set equal to 10. In all other cases, the calculated Cohen's d will be used. Also, if the reestimated sample size is lower than $f * n$, the trial will be stopped and the final analysis will be performed. For the explanation of the process of interim sample size reestimation, the reader is referred to section 2.2.

4 | RESULTS

4.1 | Power

First, it was examined how series of N-of-1 trials performed, and how they performed compared to series of N-of-1 trials with a fixed sample size, both evaluated in terms of power. Figure 1 (corresponding table A1 can be found in the appendix) displays the power of series of N-of-1 trials under various combinations of hypothesized values for the nuisance parameters (ψ_h^2 and σ_h^2) and under various values for the nuisance parameters for data generation ψ_t^2 and σ_t^2). First of all, underpowering a study appears to be least evident when 75% of the initial sample size is used to base interim sample size reestimation on. However, an excess of power appears to be problematic for some of these scenarios, especially those scenarios where $\psi_t^2 = 0.5$. Where $\psi_h^2 = 2$ it reaches even a power of 100%. Overpowering a study might be less of a problem compared to underpowering a study because the results are still reliable, it is still problematic because it is a waste of resources. Especially for studies conducted in populations of patients with rare diseases overpowering a study is not desirable.

When 50% of the initial sample size is used for reestimation of the sample size, overall power across all the scenarios appears to be adequate with some exceptions. For those scenarios where $\psi_t^2 = 2$ and $\psi_h^2 = 0.5$, and to a lesser extend also where $\psi_t^2 = 2$ and $\psi_h^2 = 1$, a lack of power occurs. However, this deficiency in power remains limited, the lowest power being 72.5%. Those scenarios where $\psi_t^2 = 0.5$ again result in an excess of power, especially those scenarios where $\psi_h^2 = 2$. It becomes clear that those scenarios where the discrepancy between the population value and the a priori expected value for ψ^2 is largest, that those scenarios have the greatest influences on the power. The value of σ_h^2 appears to have less influence on the power. Overall, it appears that taking 50% of the initial sample size has the highest chance of having an adequately powered series of N-of-1 trials.

When 25% of the initially required sample size is used to base interim sample size reestimation on, the results become less optimistic. Those scenarios where $\psi_t^2 = 2$ and $\psi_h^2 = 0.5$, regardless of the value of σ_h^2, the power drops to even (approximately) 60%. The scenarios where $\psi_h^2 = 1$ lead to adequate power only where ψ_t^2 is small. The scenarios where $\psi_h^2 = 2$ result in adequately powered series of N-of-1 trials, with the exception of the scenarios where $\psi_t^2 = 2$.

Now, the comparison of power between series of N-of-1 trials including interim sample size reestimation with series of N-of-1 trials that have a fixed sample size. Figure 2 (corresponding table A3 can be found in the appendix) shows the power for series of N-of-1 trials with a fixed sample size under various combinations of the hypothesized values for the nuisance parameters and values for the nuisance parameters used for data generation. It becomes immediately apparent that the results for series of N-of-1 trials with a fixed sample size are much more diverse and for some scenarios more problematic. Those scenarios where $\psi_h^2 = 0.5$ and $\sigma_h^2 = 0.25$ even result in studies with less than 40% power, whereas scenarios where $\psi_h^2 = 2$ will lead to studies with a power of 100%. The scenarios for the hypothesized nuisance parameters that match the population value result in adequately powered studies, as expected, but almost all the other combinations will result in over- or underpowered studies.

Compared to series of N-of-1 trials including interim sample size reestimation, the fixed sample size approach can result in fairly over- or underpowered studies. The process of interim sample size reestimation appears to be able to make up for the wrong assumptions that where made prior to the study. Take for instance the scenario where $\psi_t^2 = 0.5$ and $\sigma_t^2 = 1$, and where $\psi_h^2 = 0.5$ and $\sigma_h^2 = 0.25$, quite a moderate scenario where σ^2 is hypothesized to be lower than it is in the population. If one chooses to take the fixed sample size approach, the power of the study would be 60.8%. However, taking the two-stage approach with interim sample size reestimation, the sample size is reconsidered somewhere along the trial, and the power is restored to 83.4% when 75% of the initial sample size is used to base reestimation on, 80.4% when 50% of the initial sample size is used, and 72.7% when 25% of the initial sample size is used for reestimation. Using 25% of the initial sample size to base interim sample size reestimation might still not be optimal, but using 50% and 75% of the initial sample size results in most cases to adequately powered studies.

4.2 | Type I error rate

Next to evaluating and comparing the power in series of N-of-1 trials, the effect of interim sample size reestimation in this trial design on the type I error rate was evaluated. Figure 3 (corresponding table A3 can be found in the appendix) displays the type I error rates under the various combinations of hypothesized values for the nuisance parameters and values of the nuisance parameters for the data. Where 25% of the initially required sample size is taken to base interim sample size reestimation on, the type I error rate appears to become inflated under all the scenarios that are considered. For some scenarios, such as the combination of $\psi_t^2 = 0.25$, $\sigma_t^2 = 0.5$, $\psi_h^2 = 2$ and $\sigma_h^2 = 1$, the inflation of α remains limited. However, when 50% of the initial sample size is used as a basis for interim reestimation of the sample size, the inflation of the type I error rate appears to be

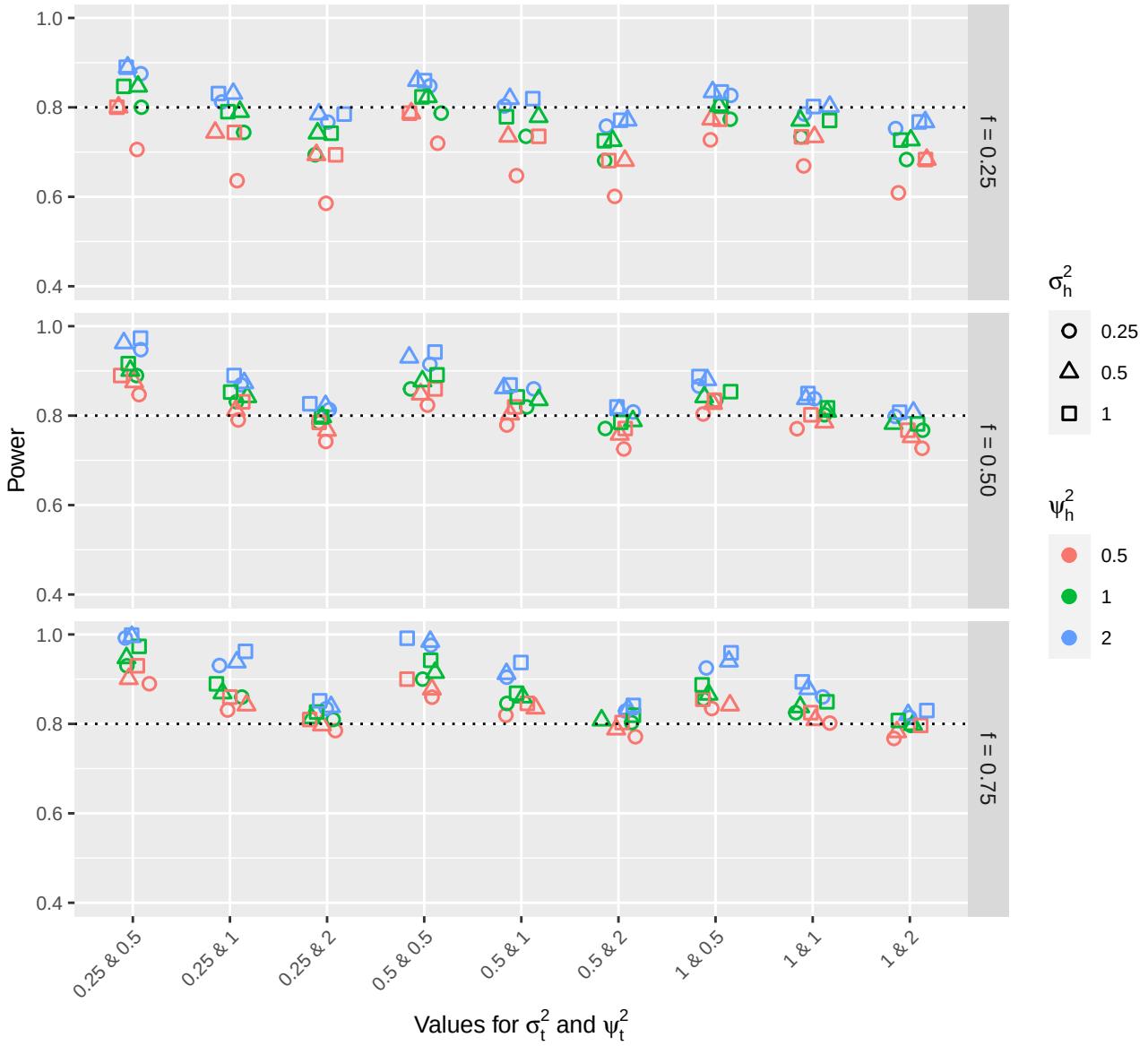

FIGURE 1 Power in simulated series of N-of-1 trials with interim sample size reestimation under different data and hypothesized values for the nuisance parameters. The x-axis indicates the different combinations of the nuisance parameters for data generation (ψ^2 and σ^2, respectively). The dotted line indicates a power of 0.8.

limited to a number of scenarios. Where ψ_h^2 is hypothesized to be equal to 2, the largest value considered in this thesis, the type I error rate appears to be closest to the nominal α level of 0.05. A larger value for σ_h^2 also appears to positively influence the type I error rate, but the value for ψ_h^2 seems to have a higher impact.

When 75% of the initial sample size is used for reestimation of the sample size at interim, it appears that the scenarios where ψ_h^2 is hypothesized to be 0.5 results in the inflation of the type I error rate. The only scenario where this does not happen, is when ψ_h^2 and σ_h^2 are both small (0.5 and 0.25 respectively) and the discrepancy between hypothesized and true values is thus smallest. For most of the scenarios where $\psi_h^2 = 2$, the type I error rate remains controlled. Where $\psi_h^2 = 1$ it appears that the type I error rate also remains controlled for the scenarios where $\psi_t^2 = 0.5$. The scenarios where $\psi_t^2 = 2$ and ψ_h^2 is hypothesized to be smaller than that seems to lead to an inflated type I error rate.

From these results it becomes clear that the value of ψ_h^2 has a greater influence on the inflation of the type I error rate than the value of σ_h^2. When the discrepancy between true and hypothesized values for ψ^2 is large, the type I error rate become (more) inflated. When one applies interim sample size reestimation in the manner that is discussed in this thesis, the total sample size is

FIGURE 2 Power in simulated series of N-of-1 trials with a fixed sample size under different data and hypothesized values for the nuisance parameters. The x-axis indicates the different combinations of the nuisance parameters for data generation (ψ^2 and σ^2, respectively). The dotted line indicates a power of 0.8.

not a constant but depends on the data that is obtained up until the interim point. A pooled estimate for the nuisance parameters is used to obtain a t statistic, which may cause the t statistic to not be t distributed in this type of design. The discrepancy between true and hypothesized values for ψ^2 appears to have the most influence on the type I error rate. If the hypothesized value for ψ^2 is small whereas the population value is large(r), the interim estimate in that case deviates a lot from the initial estimate of ψ^2. Because of the relatively big discrepancy between ψ_h^2 and $\hat{\psi}^2$, the pooled estimate deviates a lot from what it should be. If ψ_h^2 and $\hat{\psi}^2$ are close, the pooled estimate remains quite the same to what it would have been under a fixed sample size.

Taking 25% of the initial sample size for reestimation also appears to impact the inflation of the type I error rate negatively. Small sample sizes can cause the interim estimates for the nuisance parameters to be relatively unstable, thus creating a larger discrepancy between the hypothesized and interim estimates of the nuisance parameters.

5 | DISCUSSION

The series of N-of-1 trials design offers a rigorous method for minimizing the sample size in clinical trials, offering opportunities for clinical research in populations of people with rare diseases, but also for general clinical research. The required sample size is determined by a number of factors, under which the population nuisance parameters which are, to the contrary of other factors such as type I and II error rates and the clinically relevant effect, generally unknown. There is usually considerable uncertainty about the values that are hypothesized for these unknown parameters. To deal with this problem, this thesis investigated the use of interim sample size reestimation in series of N-of-1 trials to contribute to the improvements of reliable trial methodology in populations of people with rare diseases. With interim sample size reestimation, the variances are estimated during an ongoing trial and those estimates are used to recalculate the sample size to make up for misspecifications prior to the trials.

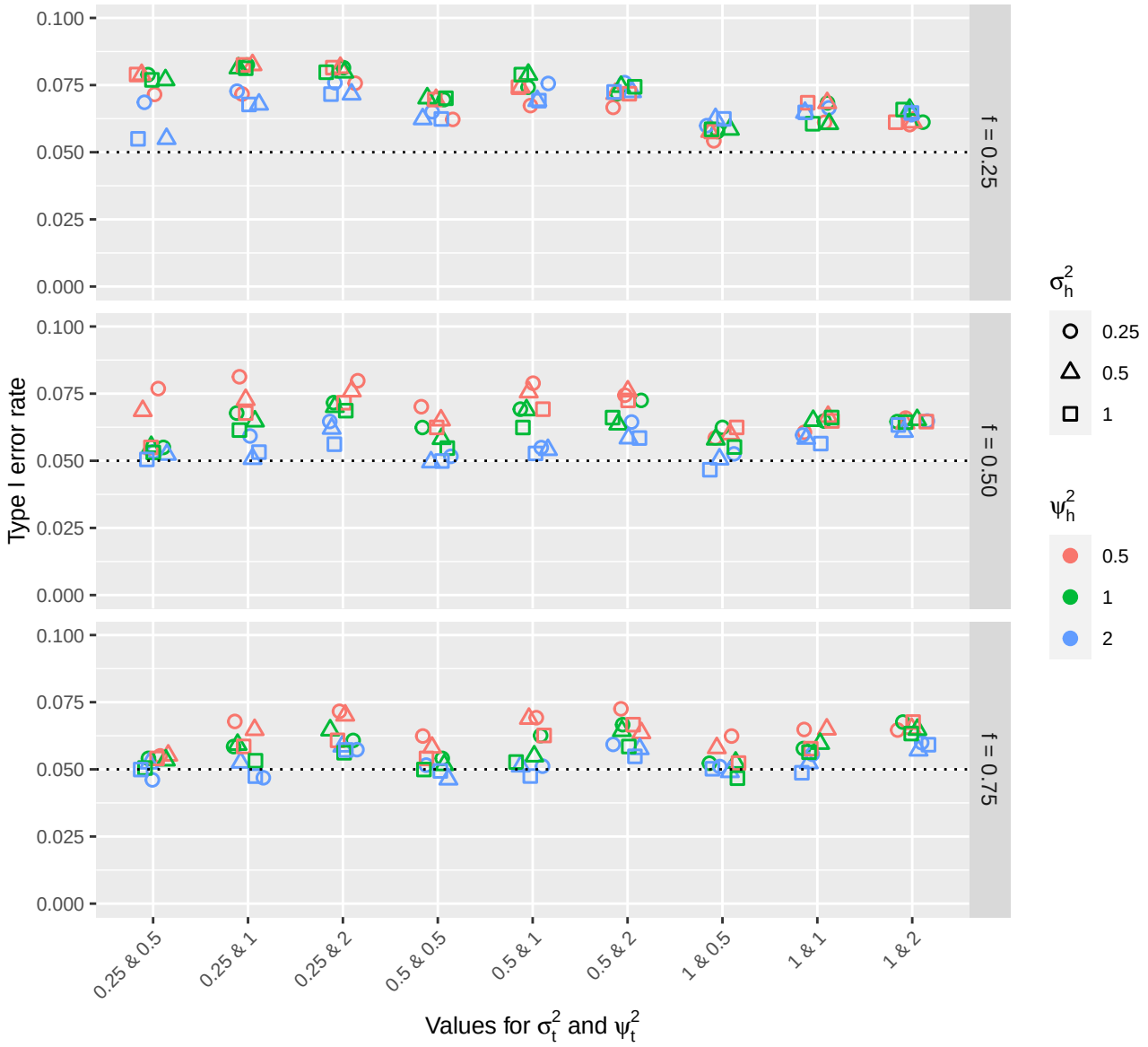

FIGURE 3 Type I error rate in simulated series of N-of-1 trials under different data and hypothesized values for the nuisance parameters. The x-axis indicates the different combinations of the nuisance parameters for data generation (ψ^2 and σ^2, respectively). The dotted line indicates the nominal α level of 0.05.

In this thesis, two simulation studies are performed to examine the reliability of series of N-of-1 trials in terms of power and the type I error rate, and to compare series of N-of-1 trials including interim sample size reestimation with series of N-of-1 trials that have a fixed sample size. Based on the results for the power and type I error rate, recommendations will be made with regard to the minimally required sample size to base interim reestimation of the sample size on.

The results indicate that series of N-of-1 trials including interim sample size reestimation have higher power than series of N-of-1 trials with a fixed sample size. Especially for those trials where the misspecification of the nuisance parameters prior to the study was large compared to the population value, interim sample size reestimation in series of N-of-1 trials has a large positive impact on the power of the study in comparison with similar fixed sample size trials. When 50% or 75% of the initial sample size is used to base interim sample size reestimation on, almost all scenarios considered in this thesis will lead to a power of the desired 80% or higher. These scenarios where the power is much higher than 80% are not desirable. However, compared to the low levels of power for some scenarios of the fixed sample size design, incorporating interim sample size reestimation in series of N-of-1 trials can be a demonstrably improvement.

The type I error probabilities in series of N-of-1 trials appeared to be inflated for most of the scenarios considered in this thesis. The method of interim sample size reestimation as discussed causes the total sample size to be dependent on the data that is obtained up until the interim point. A pooled estimate for each nuisance parameter is used, which may cause the statistic to not be distributed as expected in this approach[16]. Small initial sample sizes can cause the interim estimates of the nuisance parameters to be unreliable. The larger the discrepancy between the interim estimates for the nuisance parameters and the population values, the larger the discrepancy between the pooled estimate and what it should be. For larger initial sample sizes, the interim estimates of the nuisance parameters can be estimated with more reliability, causing higher chances for the pooled estimates to be closer to their true values.

Considering that the type I error rate is more likely to be controlled at the nominal level when the initial sample size is relatively large, or at least large enough to estimate the nuisance parameters at interim reliably, and that statistical power is also more adequate when a larger fraction of the initial sample size is used to base reestimation on, it appears to be best that the initial sample size is not too small. However, as this research focuses on improving trial designs for clinical research in small populations, it is also desirable to minimize the number of patients to include in a clinical study. For adequately powered series of N-of-1 trials, the minimum sample size for reliable interim sample size reestimation should be at least 20 patients. Then, 50% of the initial sample size can be used to base interim reestimation on. For smaller sample sizes, 75% of the initial sample size to base reestimation on will lead to adequately powered series of N-of-1 trials. Notice that it is desirable to take a smaller fraction of the initial sample size for reestimation, as chances of overpowering are higher when a larger fraction of the initial sample size is used and the misspecification of the nuisance parameters is large. The type I error rate may not be controlled at the nominal α level when 50% or even 75% of the initial sample size is used for reestimation. Only when the sample size is 20 and 75% of the initial sample size is used for reestimation, the type I error rate is controlled at the nominal level. Methods to control the bias in the type I error rate at the nominal level exist[20,16], but not specifically for the series of N-of-1 trials design. This can be an opportunity for future research.

Some limitations of this study are acknowledged. First of all, only a limited number of design scenarios are discussed here. It was attempted to make the scenarios that are considered in this study as realistic as possible so that these are applicable to real life situations. However, other designs could have different results and implications. Second, practical issues such as carryover effects and selection bias, issues that can significantly influence the results, are not considered in the simulation studies of this paper. Third, and last, this thesis does not investigate the effect of the relationship between the number of patients and the number of cycles within a patient on the sample size. As the number of cycles within a patient is increased, the number of required patients would be expected to decrease. Investigating the relationship between the number of cycles and patients on the sample size and the implications of this relationship on interim sample size reestimation could be very valuable for the development of research in clinical trial designs for populations of patients with rare diseases.

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. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly* 2004; 82(4): 661–687. doi: 10.1111/j.0887-378X.2004.00327.x

3. Johnston BC, Mills E. N-of-1 randomized controlled trials: an opportunity for complementary and alternative medicine evaluation. *Journal of Alternative & Complementary Medicine* 2004; 10(6): 979–984. doi: 10.1089/acm.2004.10.979

4. 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. doi: 10.1016/j.jclinepi.2010.05.009

5. 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. doi: 10.1016/S0895-4356(96)00429-5

6. 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. doi: 10.1001/jama.2018.18020

7. 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. doi: 10.1016/j.jpainsymman.2015.03.009

8. 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. doi: 10.7326/M18-0517

9. Senn S. Sample size considerations for n-of-1 trials. *Statistical methods in medical research* 2019; 28(2): 372–383. doi: 10.1177/0962280217726801

10. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. *PloS one* 2016; 11(12): e0167167. doi: 10.1371/journal.pone.0167167

11. Zucker DM, Denne J. Sample–size redetermination for repeated measures studies. *Biometrics* 2002; 58(3): 548–559. doi: 10.1111/j.0006-341X.2002.00548.x

12. Senn S. Mastering variation: variance components and personalised medicine. *Statistics in medicine* 2016; 35(7): 966–977. doi: 10.1002/sim.6739

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. doi: 10.1081/BIP-200062852

14. 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. doi: 10.1002/sim.4780090113

15. Proschan MA. Sample size re-estimation in clinical trials. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 2009; 51(2): 348–357. doi: 10.1002/bimj.200800266

16. 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.

17. Birkett MA, Day SJ. Internal pilot studies for estimating sample size. *Statistics in medicine* 1994; 13(23-24): 2455–2463. doi: 10.1002/sim.4780132309

18. Champely S, Ekstrom C, Dalgaard P, et al. Package 'pwr'. *R package version* 2018; 1(2).

19. R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2021.

20. Coffey CS, Muller KE. Exact test size and power of a Gaussian error linear model for an internal pilot study. *Statistics in Medicine* 1999; 18(10): 1199–1214.

21. Neyman J, Pearson ES. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 1933; 231(694-706): 289–337. doi: 10.1098/rsta.1933.0009

22. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 2015; 67(1): 1–48. doi: 10.18637/jss.v067.i01

23. Corbeil RR, Searle SR. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* 1976; 18(1): 31–38. doi: 10.2307/1267913

6 | APPENDIX

TABLE A1 Power of series of N-of-1 trials including interim sample size reestimation.

	$\psi_t^2 = 0.5; \sigma_t^2 = 0.25$			$\psi_t^2 = 1; \sigma_t^2 = 0.25$			$\psi_t^2 = 2; \sigma_t^2 = 0.25$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.706	0.847	0.890	0.636	0.790	0.831	0.585	0.742	0.785
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.800	0.890	0.930	0.744	0.831	0.860	0.694	0.785	0.810
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.875	0.948	0.992	0.812	0.869	0.931	0.767	0.813	0.834
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.800	0.875	0.901	0.744	0.812	0.842	0.694	0.767	0.797
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.847	0.901	0.948	0.790	0.842	0.869	0.742	0.797	0.813
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.890	0.963	0.995	0.831	0.873	0.938	0.785	0.822	0.838
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.800	0.890	0.930	0.744	0.831	0.860	0.694	0.785	0.810
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.847	0.916	0.973	0.790	0.853	0.890	0.742	0.797	0.827
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.890	0.973	0.999	0.831	0.890	0.962	0.785	0.827	0.852
	$\psi_t^2 = 0.5; \sigma_t^2 = 0.5$			$\psi_t^2 = 1; \sigma_t^2 = 0.5$			$\psi_t^2 = 2; \sigma_t^2 = 0.5$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.720	0.823	0.860	0.647	0.779	0.819	0.601	0.725	0.771
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.787	0.860	0.900	0.735	0.819	0.846	0.681	0.771	0.803
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.848	0.915	0.976	0.803	0.860	0.904	0.758	0.808	0.829
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.787	0.848	0.877	0.735	0.803	0.835	0.681	0.758	0.788
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.823	0.877	0.915	0.779	0.835	0.860	0.725	0.788	0.808
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.860	0.930	0.984	0.819	0.862	0.912	0.771	0.813	0.833
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.787	0.860	0.900	0.735	0.819	0.846	0.681	0.771	0.803
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.823	0.891	0.942	0.779	0.842	0.869	0.725	0.785	0.819
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.860	0.942	0.992	0.819	0.869	0.937	0.771	0.819	0.841
	$\psi_t^2 = 0.5; \sigma_t^2 = 1$			$\psi_t^2 = 1; \sigma_t^2 = 1$			$\psi_t^2 = 2; \sigma_t^2 = 1$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.727	0.804	0.834	0.669	0.771	0.802	0.609	0.727	0.767
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.773	0.834	0.855	0.734	0.802	0.825	0.683	0.767	0.797
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.827	0.866	0.925	0.785	0.838	0.860	0.753	0.798	0.815
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.773	0.827	0.842	0.734	0.785	0.809	0.683	0.753	0.782
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.804	0.842	0.866	0.771	0.809	0.838	0.727	0.782	0.798
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.834	0.880	0.940	0.802	0.838	0.878	0.767	0.808	0.820
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.773	0.834	0.855	0.734	0.802	0.825	0.683	0.767	0.797
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.804	0.854	0.887	0.771	0.817	0.849	0.727	0.782	0.808
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.834	0.887	0.959	0.802	0.849	0.894	0.767	0.808	0.830

TABLE A2 Power of series of N-of-1 trials with a fixed sample size.

	$\psi_t^2 = 0.5; \sigma_t^2 = 0.25$	$\psi_t^2 = 1; \sigma_t^2 = 0.25$	$\psi_t^2 = 2; \sigma_t^2 = 0.25$
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.836	0.611	0.379
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.973	0.822	0.558
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.999	0.974	0.815
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.892	0.674	0.427
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.981	0.864	0.611
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	1.000	0.979	0.837
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.973	0.822	0.558
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.994	0.930	0.708
$\psi_h^2 = 2; \sigma_h^2 = 1$	1.000	0.989	0.878
	$\psi_t^2 = 0.5; \sigma_t^2 = 0.5$	$\psi_t^2 = 1; \sigma_t^2 = 0.5$	$\psi_t^2 = 2; \sigma_t^2 = 0.5$
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.752	0.554	0.358
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.933	0.764	0.530
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.995	0.954	0.786
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.813	0.615	0.401
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.952	0.817	0.573
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.997	0.963	0.807
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.933	0.764	0.530
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.983	0.895	0.675
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.999	0.976	0.849
	$\psi_t^2 = 0.5; \sigma_t^2 = 1$	$\psi_t^2 = 1; \sigma_t^2 = 1$	$\psi_t^2 = 2; \sigma_t^2 = 1$
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.608	0.468	0.320
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.822	0.672	0.480
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.973	0.901	0.734
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.675	0.521	0.358
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.862	0.719	0.518
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.979	0.920	0.757
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.822	0.672	0.480
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.927	0.821	0.617
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.987	0.942	0.803

TABLE A3 Type I error rate of series of N-of-1 trials including interim sample size reestimation.

	$\psi_t^2 = 0.5; \sigma_t^2 = 0.25$			$\psi_t^2 = 1; \sigma_t^2 = 0.25$			$\psi_t^2 = 2; \sigma_t^2 = 0.25$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.071	0.077	0.055	0.072	0.081	0.068	0.076	0.080	0.072
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.079	0.055	0.054	0.083	0.068	0.059	0.082	0.072	0.061
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.069	0.053	0.046	0.073	0.059	0.047	0.076	0.065	0.057
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.079	0.069	0.055	0.083	0.073	0.065	0.082	0.076	0.070
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.077	0.055	0.053	0.081	0.065	0.059	0.080	0.070	0.065
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.055	0.052	0.052	0.068	0.051	0.053	0.072	0.062	0.059
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.079	0.055	0.054	0.083	0.068	0.059	0.082	0.072	0.061
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.077	0.053	0.051	0.081	0.061	0.053	0.080	0.069	0.056
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.055	0.051	0.050	0.068	0.053	0.048	0.072	0.056	0.057
	$\psi_t^2 = 0.5; \sigma_t^2 = 0.5$			$\psi_t^2 = 1; \sigma_t^2 = 0.5$			$\psi_t^2 = 2; \sigma_t^2 = 0.5$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.062	0.070	0.062	0.067	0.079	0.069	0.067	0.074	0.073
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.070	0.062	0.054	0.074	0.069	0.063	0.072	0.073	0.067
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.065	0.052	0.052	0.076	0.055	0.051	0.076	0.064	0.059
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.070	0.065	0.058	0.074	0.076	0.069	0.072	0.076	0.064
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.070	0.058	0.052	0.079	0.069	0.055	0.074	0.064	0.064
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.062	0.050	0.046	0.069	0.054	0.051	0.073	0.058	0.058
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.070	0.062	0.054	0.074	0.069	0.063	0.072	0.073	0.067
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.070	0.055	0.050	0.079	0.062	0.053	0.074	0.066	0.059
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.062	0.050	0.049	0.069	0.053	0.048	0.073	0.059	0.055
	$\psi_t^2 = 0.5; \sigma_t^2 = 1$			$\psi_t^2 = 1; \sigma_t^2 = 1$			$\psi_t^2 = 2; \sigma_t^2 = 1$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	0.054	0.059	0.062	0.061	0.061	0.065	0.060	0.066	0.065
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	0.058	0.062	0.052	0.068	0.065	0.058	0.061	0.065	0.068
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	0.060	0.053	0.051	0.067	0.060	0.056	0.064	0.065	0.060
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	0.058	0.060	0.058	0.068	0.067	0.065	0.061	0.064	0.065
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	0.059	0.058	0.053	0.061	0.065	0.060	0.066	0.065	0.065
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	0.062	0.051	0.049	0.065	0.058	0.052	0.065	0.061	0.057
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	0.058	0.062	0.052	0.068	0.065	0.058	0.061	0.065	0.068
$\psi_h^2 = 1; \sigma_h^2 = 1$	0.059	0.055	0.047	0.061	0.066	0.056	0.066	0.064	0.063
$\psi_h^2 = 2; \sigma_h^2 = 1$	0.062	0.047	0.050	0.065	0.056	0.049	0.065	0.063	0.059

TABLE A4 Average sample size (standard deviation in brackets) in series of N-of-1 trials after interim sample size reestimation.

	$\psi_t^2 = 0.5; \sigma_t^2 = 0.25$			$\psi_t^2 = 1; \sigma_t^2 = 0.25$			$\psi_t^2 = 2; \sigma_t^2 = 0.25$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
True sample size		**7.38**			**11.23**			**19.02**	
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	8.25 (7.20)	7.94 (4.16)	7.97 (3.25)	12.11 (12.81)	11.85 (7.15)	11.79 (5.70)	19.28 (23.26)	19.69 (13.91)	19.53 (10.73)
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	8.12 (5.10)	7.97 (3.25)	7.90 (2.58)	11.99 (9.17)	11.79 (5.70)	11.73 (4.54)	19.69 (16.76)	19.53 (10.73)	19.45 (8.46)
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	7.98 (3.64)	7.88 (2.43)	7.89 (1.93)	11.74 (6.35)	11.75 (4.33)	11.76 (3.45)	19.37 (11.70)	19.49 (8.04)	19.47 (6.36)
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	8.12 (5.10)	7.98 (3.64)	7.95 (2.95)	11.99 (9.17)	11.74 (6.35)	11.67 (5.22)	19.69 (16.76)	19.37 (11.70)	19.60 (9.78)
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	7.94 (4.16)	7.95 (2.95)	7.88 (2.43)	11.85 (7.15)	11.67 (5.22)	11.75 (4.33)	19.69 (13.91)	19.60 (8.78)	19.49 (8.04)
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	7.97 (3.25)	7.88 (2.31)	7.87 (1.90)	11.79 (5.70)	11.78 (4.08)	11.73 (3.31)	19.53 (10.73)	19.63 (7.58)	19.42 (6.12)
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	8.12 (5.10)	7.97 (3.25)	7.90 (2.58)	11.90 (9.17)	11.79 (5.70)	11.73 (4.54)	19.69 (16.76)	19.53 (10.73)	19.45 (8.46)
$\psi_h^2 = 1; \sigma_h^2 = 1$	7.94 (4.16)	7.87 (2.71)	7.85 (2.18)	11.85 (7.15)	11.84 (4.93)	11.71 (3.84)	19.69 (13.91)	19.58 (9.08)	19.63 (7.27)
$\psi_h^2 = 2; \sigma_h^2 = 1$	7.97 (3.25)	7.85 (2.18)	7.90 (1.79)	11.79 (5.70)	11.71 (3.84)	11.75 (3.14)	19.53 (10.73)	19.63 (7.27)	19.56 (5.84)
	$\psi_t^2 = 0.5; \sigma_t^2 = 0.5$			$\psi_t^2 = 1; \sigma_t^2 = 0.5$			$\psi_t^2 = 2; \sigma_t^2 = 0.5$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
True sample size		**8.65**			**12.52**			**20.32**	
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	10.07 (8.92)	9.41 (5.02)	9.33 (3.94)	13.98 (14.88)	13.18 (8.20)	13.15 (6.44)	22.22 (26.10)	20.91 (14.67)	20.72 (11.52)
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	9.59 (6.16)	9.33 (3.94)	9.20 (3.16)	13.49 (10.07)	13.15 (6.44)	13.02 (5.22)	21.31 (18.53)	20.72 (11.52)	20.83 (9.12)
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	9.40 (4.42)	9.27 (2.99)	9.15 (2.38)	13.12 (7.13)	13.03 (4.89)	13.01 (3.97)	21.09 (12.90)	20.81 (8.53)	20.83 (6.87)
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	9.59 (6.16)	9.40 (4.42)	9.25 (3.54)	13.49 (10.07)	13.12 (7.13)	13.06 (5.93)	21.31 (18.53)	21.09 (12.90)	20..85 (10.30)
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	9.41 (5.02)	9.25 (3.54)	9.27 (2.99)	13.18 (8.20)	13.06 (5.93)	13.03 (4.89)	20.91 (14.67)	20.85 (10.30)	20.81 (8.53)
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	9.33 (3.94)	9.17 (2.80)	9.17 (2.36)	13.15 (6.44)	13.04 (4.61)	13.06 (3.82)	20.72 (11.52)	20.73 (8.16)	20.75 (6.65)
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	9.59 (6.16)	9.33 (3.94)	9.20 (3.16)	13.49 (10.07)	13.15 (6.44)	13.02 (5.22)	21.31 (18.53)	20.72 (11.52)	20.83 (9.12)
$\psi_h^2 = 1; \sigma_h^2 = 1$	9.41 (5.02)	9.28 (3.38)	9.19 (2.71)	13.18 (8.20)	13.10 (5.50)	13.00 (4.29)	20.91 (14.67)	20.81 (9.65)	20.74 (7.70)
$\psi_h^2 = 2; \sigma_h^2 = 1$	9.33 (3.94)	9.19 (2.71)	9.14 (2.21)	13.15 (6.44)	13.00 (4.29)	12.98 (3.52)	20.72 (11.52)	20.74 (7.70)	20.90 (6.29)
	$\psi_t^2 = 0.5; \sigma_t^2 = 1$			$\psi_t^2 = 1; \sigma_t^2 = 1$			$\psi_t^2 = 2; \sigma_t^2 = 1$		
f	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
True sample size		**11.23**			**15.11**			**22.93**	
$\psi_h^2 = 0.5; \sigma_h^2 = 0.25$	13.70 (12.52)	12.62 (6.89)	12.23 (5.35)	17.13 (17.44)	16.32 (10.15)	15.78 (7.96)	24.59 (28.88)	24.08 (17.00)	23.57 (12.92)
$\psi_h^2 = 1; \sigma_h^2 = 0.25$	12.82 (8.43)	12.23 (5.35)	12.05 (4.26)	16.54 (12.32)	15.78 (7.96)	15.73 (6.29)	23.99 (20.32)	23.57 (12.92)	23.44 (10.33)
$\psi_h^2 = 2; \sigma_h^2 = 0.25$	12.37 (5.87)	12.01 (4.07)	11.86 (3.27)	15.89 (8.74)	15.73 (5.99)	15.61 (4.91)	23.71 (14.72)	23.52 (9.76)	23.37 (7.97)
$\psi_h^2 = 0.5; \sigma_h^2 = 0.5$	12.82 (8.43)	12.37 (5.87)	12.23 (4.99)	16.54 (12.32)	15.89 (8.74)	15.73 (7.14)	23.99 (20.32)	23.71 (14.72)	23.61 (11.94)
$\psi_h^2 = 1; \sigma_h^2 = 0.5$	12.62 (6.89)	12.23 (4.99)	12.01 (4.07)	16.32 (10.15)	15.73 (7.14)	15.73 (5.99)	24.08 (17.00)	23.61 (11.94)	23.52 (9.76)
$\psi_h^2 = 2; \sigma_h^2 = 0.5$	12.23 (5.35)	11.92 (3.84)	11.80 (3.16)	15.78 (7.96)	15.64 (5.70)	15.65 (4.73)	23.57 (12.92)	23.44 (9.11)	23.34 (7.64)
$\psi_h^2 = 0.5; \sigma_h^2 = 1$	12.82 (8.43)	12.23 (5.35)	12.05 (4.26)	16.54 (12.32)	15.78 (7.96)	15.73 (6.29)	23.99 (20.32)	23.57 (12.92)	23.44 (10.33)
$\psi_h^2 = 1; \sigma_h^2 = 1$	12.62 (6.89)	12.08 (4.44)	11.96 (3.71)	16.32 (10.15)	15.77 (6.66)	15.65 (5.49)	24.08 (17.00)	23.43 (11.10)	23.51 (8.95)
$\psi_h^2 = 2; \sigma_h^2 = 1$	12.23 (5.35)	11.96 (3.71)	11.85 (3.00)	15.78 (7.96)	15.65 (5.49)	15.60 (4.43)	23.57 (12.92)	23.51 (8.95)	23.53 (7.09)