



Published in final edited form as:

Harv Data Sci Rev. 2022 ; 2022(Spec Iss 3): . doi:10.1162/99608f92.e11adff0.

Evaluating Personalized (N-of-1) Trials in Rare Diseases: How Much Experimentation Is Enough?

Ken Cheung¹, Hiroshi Mitsumoto²

¹Mailman School of Public Health, Columbia University, New York City, New York, United States of America

²Columbia University Irving Medical Center, Columbia University, New York City, New York, United States of America

Abstract

For rare diseases, conducting large, randomized trials of new treatments can be infeasible due to limited sample size, and it may answer the wrong scientific questions due to heterogeneity of treatment effects. Personalized (N-of-1) trials are multi-period crossover studies that aim to estimate individual treatment effects, thereby identifying the optimal treatments for individuals. This article examines the statistical design issues of evaluating a personalized (N-of-1) treatment program in people with amyotrophic lateral sclerosis (ALS). We propose an evaluation framework based on an analytical model for longitudinal data observed in a personalized trial. Under this framework, we address two design parameters: length of experimentation in each trial and number of trials needed. For the former, we consider patient-centric design criteria that aim to maximize the benefits of enrolled patients. Using theoretical investigation and numerical studies, we demonstrate that, from a patient's perspective, the duration of an experimentation period should be no longer than one-third of the entire follow-up period of the trial. For the latter, we provide analytical formulae to calculate the power for testing quality improvement due to personalized trials in a randomized evaluation program and hence determine the required number of trials needed for the program. We apply our theoretical results to design an evaluation program for ALS treatments informed by pilot data and show that the length of experimentation has a small impact on power relative to other factors such as the degree of heterogeneity of treatment effects.

Keywords

ALS; heterogeneity of treatment effects (HTE); minimally clinically important heterogeneity; patient-centered research; rare diseases; sample size formulae

This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Corresponding Author: Ken Cheung, Department of Biostatistics, Columbia University Irving Medical Center, 622 W. 168th St., New York, NY 10032, yc632@cumc.columbia.edu, T: (212)-305-3332.

1. Introduction

When managing chronic diseases and conditions, patients commonly try different treatments over time before finding the right treatments. The practice of N-of-1 trials operationalizes this type of patient-centered experimentation by randomizing treatments to single patients in multiple crossover periods, often in a balanced fashion. N-of-1 trials can be used to identify the optimal personalized treatment for single patients in situations involving evidence for heterogeneity of treatment effects (HTE) or the lack of a cure (Davidson et al. 2021).

As such, these trials are sometimes called single-patient trials or personalized trials. First introduced by (Hogben and Sim 1953), N-of-1 trials have recently been applied to treat rare diseases (Roustit et al. 2018), as well as common chronic conditions such as hypertension (Kronish et al. 2019)(Samuel et al. 2019). The use of personalized (N-of-1) trials in treating rare diseases is particularly appealing because demonstrating comparative effectiveness of treatments at the population level via parallel-group randomized trials is often infeasible.

In this article, we consider personalized (N-of-1) trials of treatments for people with amyotrophic lateral sclerosis (ALS). ALS is a rare neurodegenerative disease that affects motor neurons in the brain and spinal cord. Despite the fact that two modestly effective disease-modifying medications have been approved for the treatment of ALS (Edaravone [MCI-186] ALS 19 Study Group 2017), the disease has no cure, and thus, symptomatic treatments remain an important strategy to improve the quality of life in people with ALS (Mitumoto, Brooks, and Silani 2014). In particular, muscle cramps are disabling symptoms affecting over 90% of ALS patients, with demonstrated between-patient variability and yet stable manifestation of symptoms in a patient (Caress et al. 2016). Several treatments targeting muscle cramps have been evaluated and have shown mixed results, suggesting the presence of HTE or inadequate statistical power for definitive conclusions (Baldinger, Katzberg, and Weber 2012). Furthermore, ALS itself has been considered markedly heterogeneous in its pathogenesis, disease manifestations, and disease progression (Al-Chalabi and Hardiman 2013)(van den Berg et al. 2019). These are the clinical situations in which personalized (N-of-1) trials can help patients identify the best treatments for themselves (n.d.a).

Despite renewed interest in N-of-1 trials and numerous recent applications, the literature has offered little discussion on the evaluation of the usefulness of N-of-1 trials. As N-of-1 trials typically require active physician involvement, intense monitoring, and frequent data collection compared with usual care, these additional costs and resources warrant careful evaluation of effectiveness before said trials are used in practice as regular clinical service. The primary evaluation question is “Does the practice of N-of-1 trials in clinical care improve outcomes on the standard of care?” However, presuming the quality of treatment decisions based on N-of-1 trials is higher than what standard of care would prescribe, reports of N-of-1 trials often describe only the applications and results of the trials without plans to address the evaluation question. An exception is (Kravitz et al. 2018) who compare N-of-1 intervention against the usual care for patients with musculoskeletal pain in a randomized fashion using data collected after experimentation ends and find no evidence of superior outcomes among participants undergoing N-of-1 trials. However, when planning the study, the authors had not considered the underlying model that accounts for variability and

correlation in the longitudinal observations and the assumptions on the effect size, which would in turn drive the appropriate sample size of an evaluation program for N-of-1 trials. A design issue related to sample size determination is the duration of experimentation in N-of-1 trials. In this article, we propose a framework to evaluate the quality and effectiveness of N-of-1 trials and develop specific guidance to address these design issues. We will introduce the evaluation framework in Section 2 and define the basic analytical model for analyzing N-of-1 trials in Section 3. The main findings on the experimentation duration and sample size are derived and described in Section 4 and applied to the ALS treatment program in Section 5. The article ends with a discussion in Section 6. All technical details are provided in the Appendices.

2. An Evaluation Framework for Personalized (N-of-1) Trials

2.1. The Anatomy of a Personalized (N-of-1) Trial

We consider an evaluation program comparing the effectiveness of personalized (N-of-1) trials in treating muscle cramps in people with ALS relative to the institutional standard of care. Under the program, people with ALS will be randomized to receive personalized (N-of-1) trials that compare two standard drugs prescribed for muscle cramps, mexiletine and baclofen. In each trial, a patient will be given the two drugs sequentially over $T = 18$ two-week treatment periods in two phases. The first phase consists of m treatment periods (with $m < T$) when the two drugs are randomized in a multiple crossover fashion. This phase shall be referred to as the experimentation phase. In the remaining $T - m$ treatment periods, the patient will continue with a drug treatment selected based on data in the experimentation phase. This phase shall be referred to as the validation phase (Figure 1).

During the treatment periods, the Columbia Muscle Cramp Scale (MCS) will be collected weekly to result in two MCS measurements for each period: one at the end of week 1 and one at the end of week 2. The MCS is a validated, composite score summarizing the frequency, severity, and clinical relevance of cramps in people with ALS (Mitumoto et al. 2019). While the study does not include washout periods between treatments, only the measurement at the end of each two-week period will be used in the primary analysis in order to avoid carryover effects of the drugs.

Sandwiched between the two treatment phases is a feedback period where the MCS data in the experimentation phase are reviewed with the treating physician and the patient. The feedback period enables data-driven treatment decisions by providing the stakeholders with data visualization as well as numerical comparison (Davidson et al. 2021).

2.2. Standard of Care

In this article, we focus on a randomized controlled evaluation program where patients are randomized between an N-of-1 trial and standard of care (SOC). As depicted in Figure 1, a patient under SOC will be given either mexiletine or baclofen for 36 weeks, corresponding to the 18 two-week treatment periods in the N-of-1 trials, and will have the same follow-up schedule as the N-of-1 trial patients. Treatments in the ‘experimentation phase’ will be determined by the treating physicians. The ‘feedback period’ in the SOC arm may be viewed

as a sham intervention and be conducted as a regular clinic visit before the patient continues into the ‘validation phase’ with the same drug in the remaining $T - m$ treatment periods. By the virtue of randomization, MCS collected in the validation phase under SOC will serve as the control data and allow for an unbiased comparison with the validation phase in the N-of-1 trial patients.

Let p_0 denote the probability that mexiletine will be prescribed under SOC and p_1 the probability baclofen will be prescribed such that $p_0 + p_1 = 1$. The special case $p_0 = 1$ and $p_1 = 0$ corresponds to a clinical scenario where mexiletine is considered the standard treatment. Generally, the program probability parameters p_0, p_1 are somewhere between 0 and 1 when no clear best treatment exists. A program equipoise may be defined as when the treating physicians will give either of the drugs with equal likelihood, that is, $p_0 = p_1 = 0.5$. These program parameters apparently affect the quality of treatment under standard of care, and hence the advantage of N-of-1 trials over standard of care. At the end of the evaluation program, these parameters can be estimated using the control data.

2.3. Design Parameters

While the study duration (or the number of treatment periods T) is determined based on feasibility and how long a patient can be followed in the evaluation program, an N-of-1 trial under the evaluation framework is defined by the length m of the experimentation phase, and hence the length $T - m$ of the validation phase. Intuitively, the quality of the treatment decision by an N-of-1 trial improves with a larger m as more data will be available during the feedback period. On the other hand, a long experimentation phase may place excessive burden on patients without benefitting them and imply a short validation phase for a given T . Rather than maximize accuracy, the experimentation length m will respond to the question “How much experimentation is needed for an N-of-1 trial to be beneficial to an individual?”

A second design parameter is the specification of an analytical plan used to guide treatment selection during the feedback period. Principled statistical or data science methods should be employed to ensure the analysis is rigorous, while a prespecified plan entails preprogrammed algorithms that in turn facilitate quick feedback to the stakeholders.

Finally, as in conventional randomized controlled trials, the number of patients randomized in an evaluation program will need to be determined to ensure adequate statistical power for the primary evaluation question on whether N-of-1 trials improve outcomes.

To summarize, the design parameters that need to be prespecified at the planning stage of an evaluation program are the primary analysis plan used in the feedback period, the experimentation length (m) for each individual, and the number of individuals required. These will be discussed in next two sections.

3. An Analytical Model for N-of-1 Trials

Let y_{it} be the outcome of patient i in treatment period t and $x_{it} \in \{-1, 1\}$ be the corresponding treatment for $i = 1, \dots, n$ and $t = 1, \dots, T$. Without loss of generality, we assume

a large value of the outcome y_{it} is desirable. To put the notation in the context of our study, we let y_{it} denote the negative value of MCS at the end of each two-week treatment period. For the treatments, baclofen is coded as $x_{it} = 1$ and mexiletine as $x_{it} = -1$. In this article, we focus on balanced sequences between baclofen and mexiletine in the experimentation phase, that is, assuming

$$\sum_{t=1}^m x_{it} = 0. \quad (3.1)$$

Consider the outcome model

$$y_{it} = \alpha_i + \beta_i x_{it} + \epsilon_{it} \quad (3.2)$$

where β_i is the patient-specific treatment effect and the noise ϵ_{it} s are mean zero normal with $\text{cov}(\epsilon_{it}, \epsilon_{is}) = \rho_{it} \sigma^2$ and $\rho_{it} = 1$. To reflect heterogeneous symptoms and HTE among the patients, we postulate $\alpha_i \sim N(\mu_A, \sigma_A^2)$ and $\beta_i \sim N(\mu_B, \sigma_B^2)$. The mean μ_B indicates the average treatment effect and the variance σ_B^2 indicates the extent of HTE in the disease population. While $\mu_B = 0$ represents the null scenarios where there is no average treatment effects, a large value of σ_B^2 indicates the needs for personalizing treatments.

Under model (3.2), the optimal treatment for patient i can be expressed as $2I(\beta_i > 0) - 1$, where $I(\cdot)$ is an indicator function. During the feedback period, we may present to patient i an estimated treatment effect $\hat{\beta}_i$ based on the experimentation phase data $\{(x_{it}, y_{it}) : t = 1, \dots, m\}$ along with the estimated optimal personalized treatment for the patient:

$$x_i^* = 2I(\hat{\beta}_i > 0) - 1. \quad (3.3)$$

Subsequently, in the event of perfect adherence to analysis result, the patient will receive the estimated optimal treatment (3.3) in the validation phase, that is, $x_{it} \equiv x_i^*$ for $t = m + 1, \dots, T$.

Some practical notes on the choice of $\hat{\beta}_i$ are in order. For the purposes of providing quick feedback, a broad range of estimators can be considered. The theoretical results derived in the following sections will hold for any estimators that are approximately normally distributed with mean β_i and some finite variance τ_i^2 . A simple example is the patient-specific least squares estimator $\hat{\beta}_i^{LS} = \sum_{t=1}^m x_{it} y_{it} / m$ for patient i . The least squares estimator is unbiased for the patient-specific treatment effect β_i regardless of the variance-covariance structure of $\{\epsilon\}$ with variance

$$\tau_i^2 = \text{var}\left(\widehat{\beta}_i^{LS} \mid \alpha_i, \beta_i\right) = \lambda_i \sigma^2 / m \text{ where } \lambda_i = 1 + \sum_{s \neq i} x_{is} x_{is} \rho_{st} / m. \quad (3.4)$$

Note that the conditional variance (3.4) is free of the patient-specific parameters α_i and β_i . For the purposes of planning an N-of-1 trial, we will focus on the use of least squares. However, in the actual analysis, if additional information is available to inform the appropriate correlation structure of the data, likelihood-based estimation or weighted least squares accounting for such structure may improve efficiency.

4. How Much Experimentation Is Enough?

4.1. Patient-Centric Criteria and Length of Experimentation Phase

In this subsection, we discuss the choice of the experimentation length m of an N-of-1 trial with respect to two different patient-centric criteria, both of which aim to maximize the benefits to patients on N-of-1 trials.

The first criterion is defined as the expected number of periods where a patient receives the optimal treatment. Mathematically, this criterion is denoted as $E(z_i)$, where z_i is the number of periods in which patient i receives the optimal treatment over the T treatment periods.

Proposition 1.—*Suppose $\widehat{\beta}_i \sim N(\beta_i, \tau_i^2)$ under a balanced experimentation phase (3.1). Then for $0 < m \leq T$,*

$$E(z_i) = \frac{m}{2} + (T - m) \Pr\left(W \leq \frac{|\mu_B + \sigma_B U|}{\tau_i}\right)$$

where W, U are independent standard normal variables. Furthermore, if $\mu_B = 0$, then

$$E(z_i) = \frac{m}{2} + (T - m) G(\sigma_B / \tau_i) \quad (4.1)$$

where G is the cumulative distribution function of $W/|U|$, which is a pivotal distribution.

The second patient-centric criterion is defined as the expected average outcome of a patient during an N-of-1 trial. This criterion is denoted as $E(\bar{y}_i)$, where $\bar{y}_i = \sum_{t=1}^T y_{it} / T$ is the average outcome of the patient in all T treatment periods.

Proposition 2.—*Under the same condition as in Proposition 1, for $0 < m \leq T$,*

$$E(\bar{y}_i) = \mu_A + \left(1 - \frac{m}{T}\right) \left[\mu_B \left(2\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - 1 \right) + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) \right]$$

where Φ and ϕ respectively denote the standard normal distribution function and density.

We can derive a few practical principles from Proposition 1 and Proposition 2. First, conducting an N-of-1 trial with an experimentation length $m < T$ is generally beneficial for the patient compared to experimentation in all T period. Specifically, we can derive from Proposition 1 that the patient will receive at least half of the time, that is, $E(z_i) \geq T/2$ for all m , and attain the minimum when $m = T$. Analogously from Proposition 2, the expected average outcome will be no smaller than the population average, that is, $E(\bar{y}_i) \geq \mu_A$ for all m , and equality holds when $m = T$.

Second, we can derive from the propositions that $E(z_i)$ and $E(\bar{y}_i)$ are increasing in σ_B under the null $\mu_B = 0$. In other words, an N-of-1 trial becomes more beneficial to the patient when there is a larger variability in the treatment effects across patients.

Third, and importantly, considering the null case where $\mu_B = 0$ and when the least squares $\hat{\beta}_i^{LS}$ is used to make inference based on the experimentation phase data is instructive. Under these conditions, we can derive from Proposition 2 that the criterion $E(\bar{y}_i)$ is maximized at

$$m^* = \frac{2T}{\sqrt{9 + 8\xi_i T} + 3} \quad (4.2)$$

where $\xi_i = \frac{\sigma_B^2}{\lambda_i \sigma^2}$ and λ_i is defined in (3.4). While Equation (4.2) gives the optimal length m^* as a function σ_B , σ , and λ_i , it provides some general guidance:

Main Result 1.—The optimal experimentation length m^* is less than one-third of the total N-of-1 trial duration from a patient's perspective, that is, $m^* \lesssim T/3$.

4.2. Sample Size

In this subsection, we discuss how much experimentation is adequate in terms of the sample size enrolled to the evaluation program. We first define the quality of an N-of-1 trial as the expected health outcome under the estimated optimal treatment x_i^* (3.3). Assuming perfect adherence to the analysis results in the feedback period, the quality of an N-of-1 trial can be defined as $E(y_i^*)$ where $y_i^* = \sum_{t=m+1}^T y_{it} / (T - m)$ and the expectation is taken with respect to the distributions of α_i , β_i , x_i^* , and $\{\epsilon_{it}\}$. Analogously, we can define the quality of standard of care as $E(y_i')$ where y_i' is the average outcome observed in the validation phase for a patient under SOC and the expectation is taken under the assumption that the treatment in the experimentation phase continues to the validation phase. The primary objective of the evaluation program is to compare the quality of an N-of-1 trial and the quality of SOC. This can be formulated into a hypothesis testing problem with

$$H_0: \Delta = E(y_i^*) - E(y_i') \leq 0 \text{ versus } H_1: \Delta > 0 \quad (4.3)$$

where Δ measures the degree of quality improvement due to N-of-1 trials defined over the patient population. The hypotheses (4.3) can in turn be tested using the regular Z -statistic:

$$Z = \frac{\sqrt{n}(\bar{y}^* - \bar{y}')}{\sqrt{v^* + v'}} \quad (4.4)$$

where \bar{y}^* and v^* are respectively the sample mean and the sample variance of $\{y_i^*\}$ in the n patients randomized to an N-of-1 trial and \bar{y}' and v' are the sample mean and sample variance of $\{y_i'\}$ in the n patients randomized to SOC. Using standard arguments gives the power of the Z -test

$$\Pr\left(Z > c_\alpha | \Delta\right) \approx \Phi\left(\frac{\sqrt{n}\Delta}{\sqrt{\text{var}(y_i^*) + \text{var}(y_i')}} - c_\alpha\right) \quad (4.5)$$

where c_α is the upper α th percentile of standard normal. In Appendix C, we derive the expressions for Δ , $\text{var}(y_i^*)$, and $\text{var}(y_i')$ in (4.5) under the condition that $\tau_i^2 \equiv \tau^2$ for all i . This condition is met when the N-of-1 trial patients receive the same sequence x_{it} in the experimentation phase or when $\{\epsilon_{it}\}$ has a specific variance-covariance structure. For example, under a compound symmetry, that is, $\rho_{st} \equiv \rho$, we can show $\lambda_i \equiv \lambda = 1 - \rho$, that is, having $\tau_i^2 \equiv (1 - \rho)\sigma^2/m$ for all i . Specifically, under the assumption that the SOC treatment x_i' for a given patient is independent of the patient-specific treatment effect β_i , we have

$$\Delta = 2\mu_B \left\{ \Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - p_1 \right\} + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right), \quad (4.6)$$

$$\text{var}(y_i^*) = \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \left\{ \Delta + \mu_B(2p_1 - 1) \right\}^2 + \frac{\sigma^2}{T - m}, \quad (4.7)$$

and

$$\text{var}(y_i') = \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \mu_B^2(2p_1 - 1)^2 + \frac{\sigma^2}{T - m}. \quad (4.8)$$

The above expressions account for population-level information about the treatments through the program parameter p_1 . For example, if emerging evidence in the literature suggests slight advantage of $x_i = 1$ over $x_i = -1$, we may assume the physicians in the program will select $x_i' = 1$ with $p_1 > 0.5$. In Appendix C, we provide expressions analogous

to (4.6)–(4.8) for the situations where the physician may prescribe treatment x_i' with patient-specific knowledge in addition to the population-level parameter p_i . However, we note that using expressions (4.6)–(4.8) may adequately reflect the standard of care where treatments are chosen based on population-level information rather than patient-specific knowledge. Furthermore, under the independence assumption of x_i' and β_i , the power expression depends only on model parameters $(\sigma_A, \sigma_B, \mu_B, \sigma, \tau_i)$ for which information may be available to provide preliminary estimates and the known design parameters (p_i, m, n, T) . Finally, under the null case $\mu_B = 0$:

Main Result 2.—All else being equal, the power to demonstrate quality improvement due to N-of-1 trials (vs SOC) increases as heterogeneity of treatment effects σ_B^2 increases.

5. Numerical Illustrations: Application to ALS Patients

5.1. Optimal Length of Experimentation

We use the MCS natural history data in (Mitsumoto et al. 2019) to inform the design of the evaluation program for people with ALS. Specifically, we fitted a random effects model to the data and obtained an estimate of $\sigma_A = 4.8$ and $\sigma = 1.6$. For simplicity in illustration, we further assume that the within-subject noise is conditionally independent given the population-level parameters. Figure 2 plots the patient-centric criteria against different values of μ_B, σ_B , and m for $T = 18$. While the two criteria adopt different metrics, they are maximized when m is relatively small. In Figure 2 and in all (μ_B, σ_B) that we have considered (not shown here), the optimal values of m range from 2 to 6 for both criteria. This is consistent with what Main Result 1 implies: $m^* \lesssim T/3 = 6$.

5.2. Sample Size and Effect Size

Main Result 2 implies that σ_B^2 may be viewed as an effect size in power calculation, while the power also depends on other model parameters and design parameters. As in conventional practice, the choice of an effect size should be based on a clinically meaningful difference, whereas the other model parameters (e.g., σ_A, σ , etc) may be based on pilot data if available. Figure 3 plots the power against (n, m) for three different effect sizes σ_B for a one-sided test at 5% significance. Under each effect size, we identify the smallest n that achieves 80% power for any m and obtain that the required (n, m) are (210, 12), (60, 6), and (34, 4) respectively for $\sigma_B = 1.6, 3.2$, and 4.8. We note that under a small effect size $\sigma_B = 1.6$, the required $m = 12$ is greater than $T/3$. In light of Main Result 1, we may instead adopt $(n, m) = (210, 6)$ in order to maximize the benefits of the N-of-1 trials to the patients. The power of this modified design is 78%, which is slightly lower than the target 80%. Generally, we observe from Figure 3 that the impact of m on power is relatively small compared to that of n and σ_B except when the effect size is small ($\sigma_B = 1.6$).

To determine if a specific value of σ_B corresponds to a clinically meaningful effect size, relating σ_B to Δ using (4.6) may be useful, as Δ lives on the same scale as the measurement outcomes. In our application, a 3- to 4-point change on the MCS will represent a clinical meaningful shift. Based on the pilot data and assumptions, the effect size $\sigma_B = 1.6, 3.2, 4.8$

translate to a degree of quality improvement $\Delta = 1.2, 2.5, 3.8$ respectively. Thus, we set the sample size for this evaluation program at $n = 34$ with four treatment periods (two on mexiletine and two on baclofen) based on the results for $\sigma_B = 4.8$. Generally, the minimally clinically important heterogeneity (MCIH $\sigma_{B,\min}$) may be determined relating to the minimally clinically important change (MCID, Δ_{\min}) using (4.6).

5.3. Power for Comparing to Fully Informed SOC

The calculations in the previous subsection assume the null case $\mu_B = 0$ under which the power (4.5) does not depend on the parameter p_1 . Under a non-null case, the value of p_1 reflects how informed the practice is about the population-level treatment effect. For example, if evidence in the literature suggests $\mu_B > 0$, an informed practice will prescribe $x_i' = 1$ with $p_1 > 0.5$. Specifically, standard of care that is fully informed by the literature may correspond to $p_1 = \Pr(\beta_i > 0) = \Phi(\mu_B/\sigma_B)$.

Table 1 shows that as the average treatment effect μ_B grows larger and a fully informed SOC practice prescribes $x_i' = 1$ more often (i.e., larger p_1), the power to demonstrate quality improvement in (4.3) becomes smaller. On the one hand, this suggests that if there is overwhelming evidence favoring $x_i' = 1$ over $x_i' = -1$ in the literature, conducting N-of-1 trials will have diminished effect *provided that* the standard of care is fully informed. On the other hand, even with a large average treatment effect $\mu_B = 2.4 = 0.5\sigma_B$, quality improvement due to N-of-1 trials $\Delta > 3$, which is still clinically meaningful and the power is still reasonable high (68%) in this sensitivity analysis. This suggests that evaluating N-of-1 trials is a worthwhile endeavor unless there is overwhelming evidence of a large average treatment effect.

The numerical results in this article, and power and the patient-centric criteria in general, can be computed using tools are available at: <https://roadmap2health.io/hdsr/n1power/>.

6. Discussion

N-of-1 trials have been increasingly used as a design tool to bridge practice and science in rare diseases (Müller et al. 2021)(n.d.b). However, the literature is missing concrete guidelines on N-of-1 designs as to how much experimentation is appropriate. A fundamental issue is the articulation of a framework that will facilitate the evaluation of the usefulness of N-of-1 trials. In this article, we introduce an evaluation framework and outline the basic elements in an evaluation program for N-of-1 trials—namely, an experimentation phase, a feedback period, and a validation phase. In the literature, the reporting of N-of-1 trials mostly focuses only on the results of the experimentation phase, where patients explore the different treatments sequentially under a rigorous clinical protocol such as randomization, blinding, and scheduled follow-up. The feedback period and the validation phase are the critical elements in the planning and the conduct of N-of-1 trials but are, unfortunately, often omitted in the description of the design and the analytical plan.

Specifically, the length of the validation phase, relative to that of the experimentation phase, should be given careful consideration. We have demonstrated theoretically and numerically

that the optimal length of experimentation from the patient's perspective should be no greater than one-third of the entire study duration. This implies a relatively long validation phase, suggesting the importance of reproducing the quality of the decisions due to N-of-1 trials with additional follow-up. Our theoretical results also provide guidance on how many patients are needed in order to adequately power for testing quality improvement. Importantly, the relative length of experimentation and validation has minimal impact on the power. In other words, little conflict exists between the goal of maximizing patient benefits and maximizing power.

The feedback period facilitates evidence-based treatment decisions using data measured in the experimentation phase. Summarizing the relative benefits of the treatments via a single numerical statistic is a pragmatic way to present such evidence, because the information can be objectively presented and quickly digested by stakeholders. We have developed design calculus based on the model-based least squares estimation, which is quick to compute and produces unbiased estimates of patient-specific treatment effects under a broad range of scenarios. Other more sophisticated model-based methods may be used to deal with the more complex situations. For example, when we observe high volume of outcome measures via wearable devices, we could extend model (3.2) to an autoregressive model with multiple observations per treatment period (Kronish et al. 2019). In practice, treatment decisions are likely determined based on the totality of evidence. For example, in situations where a treatment that apparently benefits a patient may have side effects, a possibly less effective treatment may be preferred if it is more easily tolerated. Considerations of multiple outcomes in the analysis during the feedback period will likely increase adherence and will warrant further empirical, domain-specific research. Overall, as the feedback period potentially changes the treatment decisions—and hence, the outcomes—in the validation phase, it can be viewed as an integral part of the intervention component. We may thus experiment in a randomized fashion different elements in the feedback period for different individuals: we may consider presenting different endpoints (e.g., muscle cramp or safety), using a single endpoint, a composite outcome, or multivariate endpoints, using different types of analyses (e.g., intent-to-treat vs per-protocol), and asking patients for their satisfaction and preference (Cheung et al. 2020).

Some considerations, assumptions, and limitations for power calculation in conventional randomized controlled trials also apply for N-of-1 trials. First, power calculation involves the inputs of a number of nuisance model parameters (e.g., σ_A, σ) as well as the effect size (σ_B^2). While the effect size σ_B^2 should be determined based on clinical relevance shift, the other parameters ideally can be based on estimates from pilot data. However, in situations where robust pilot data are not available, a potential useful strategy is to leverage the concepts of adaptive designs (U.S. Food & Drug Administration 2019) whereby the model parameters are updated using interim data in the evaluation program and the updates in turn inform a reassessment of the degree of quality improvement and the sample size required.

Second, our derivations assume that patients in both arms comply with their treatments in the following sense: patients in the N-of-1 trials adhere to the estimated optimal treatments based on the experimentation phase data, and patients in the SOC continue with the same treatment as in the experimentation phase. If there is prior information about

noncompliance rate, power expressions can be derived accordingly under the proposed framework. However, from the viewpoint that the feedback period is part of the N-of-1 trial intervention, it should be designed to maximize adherence by choosing the outcomes and analyses that most reflect patient preference as discussed in the previous paragraph. Third, approaches to deal with missing data should be prespecified and implemented during the feedback period. An advantage of using model-based estimation is that the model can also serve as the basis for multiple imputations. That being said, no statistical approach can replace a well-conducted trial that is characterized by good compliance to treatment and minimal missing data.

Disclosure Statement

This work was supported by grants R01LM012836 from the NIH/NLM, P30AG063786 from the NIH/NIA, UL1TR001873 from NIH/NCATS, and R01MH109496 from NIH/NIMH. Dr. Mitumoto's work was also supported by ALS Association, MDA Wings Over Wall Street, Spastic Paraplegia Foundation, Mitsubishi-Tanabe, and Tsumura. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The views expressed in this paper are those of the authors and do not represent the views of the National Institutes of Health, the U.S. Department of Health and Human Services, or any other government entity.

Appendices

Appendix A. Theoretical Results Concerning $E(z_i)$

A.1. Proof of Proposition 1.

First, consider the case $\beta_i > 0$ for patient i so the optimal treatment is $x_{it} = 1$. Then, the number of periods the patient is on the optimal treatment equals

$$z_i = \frac{m}{2} + (T - m)I(x_i^* = 1) = \frac{m}{2} + (T - m)I(\hat{\beta}_i > 0). \quad (\text{A.1})$$

The first term in the right-hand-side of (A.1) is the number of optimal treatment periods received in the experimental phase, and the second term is the number in the validation phase. Since $\hat{\beta}_i \sim N(\beta_i, \tau_i^2)$, we have

$$E\{I(\hat{\beta}_i > 0) | \alpha_i, \beta_i\} = \Pr(\hat{\beta}_i > 0 | \alpha_i, \beta_i) = \Phi(\beta_i / \tau_i),$$

and therefore,

$$E(z_i | \alpha_i, \beta_i) = \frac{m}{2} + (T - m)\Phi(\beta_i / \tau_i) \text{ when } \beta_i > 0. \quad (\text{A.2})$$

Next, under the case $\beta_i < 0$, we can analogously derive that

$$E(z_i | \alpha_i, \beta_i) = \frac{m}{2} + (T - m)\Phi(-\beta_i/\tau_i) \text{ when } \beta_i < 0. \tag{A.3}$$

Combining (A.2) and (A.3) gives

$$E(z_i | \alpha_i, \beta_i) = E(z_i | \beta_i) = \frac{m}{2} + (T - m)\Phi(|\beta_i|/\tau_i), \tag{A.4}$$

which is free of α_i . Since $E(z_i) = E\{E(z_i | \beta_i)\}$, the expectations of both sides in (A.4) are to be taken with respect to the distribution of $\beta_i \sim N(\mu_B, \sigma_B^2)$ to complete the proof. By change of variable, we have

$$\begin{aligned} E\Phi(|\beta_i|/\tau_i) &= \int_{-\infty}^{\infty} \int_{-\infty}^{|\beta|/\tau_i} \frac{1}{\sigma_B} \phi(w) \Phi\left(\frac{b - \mu_B}{\sigma_B}\right) dw db \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{|\mu_B + \sigma_B u|/\tau_i} \phi(w) \Phi(u) dw du \\ &= \Pr\left(W \leq \frac{|\mu_B + \sigma_B U|}{\tau_i}\right). \end{aligned} \tag{A.5}$$

The proof is completed by substituting (A.5) into (A.4).

Appendix B. Theoretical Results Concerning $E(\bar{y}_i)$

B.1. Lemma 1

Derivations of $E(\bar{y}_i)$ will be facilitated by first noting the following lemma:

Lemma 1.

Let $V \sim N(\mu_V, \sigma_V^2)$. Then,

$$E\left[V\Phi(V)\right] = \mu_V\Phi\left(\frac{\mu_V}{\sqrt{\sigma_V^2 + 1}}\right) + \frac{\sigma_V^2}{\sqrt{\sigma_V^2 + 1}}\phi\left(\frac{\mu_V}{\sqrt{\sigma_V^2 + 1}}\right)$$

where Φ and ϕ respectively denote the standard normal distribution function and density function.

Proof of Lemma 1:

Using definition of expectation, we derive

$$\begin{aligned}
E\left\{V\Phi(V)\right\} &= \int_{-\infty}^{\infty} \int_{-\infty}^v v\phi(u) \frac{1}{\sigma_V} \phi\left(\frac{v-\mu_V}{\sigma_V}\right) dudv \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\mu_V + \sigma_V w} (\mu_V + \sigma_V w)\phi(u)\phi(w) dudw \\
&= \mu_V \Pr\left(U < \mu_V + \sigma_V W\right) + \sigma_V \int_{-\infty}^{\infty} w\Phi\left(\mu_V + \sigma_V w\right)\phi(w) dw \\
&= \mu_V \Phi\left(\frac{\mu_V}{\sqrt{1 + \sigma_V^2}}\right) + \sigma_V^2 \int_{-\infty}^{\infty} \phi(w)\phi\left(\mu_V + \sigma_V w\right) dw
\end{aligned}
\tag{B.1}$$

where U, W are independent standard normal variables. Thus, $U - \sigma_V W \sim N(0, 1 + \sigma_V^2)$, and the first term in (B.1) can be evaluated as

$$\mu_V \Pr\left(U < \mu_V + \sigma_V W\right) = \mu_V \Phi\left(\frac{\mu_V}{\sqrt{1 + \sigma_V^2}}\right).
\tag{B.2}$$

Next, the single integral in the second term in (B.2) can be evaluated using integration by parts

$$\begin{aligned}
\int_{-\infty}^{\infty} w\Phi\left(\mu_V + \sigma_V w\right)\phi(w) dw &= \sigma_V \int_{-\infty}^{\infty} \phi\left(\mu_V + \sigma_V w\right)\phi(w) dw \\
&= \sigma_V \frac{1}{\sqrt{\sigma_V^2 + 1}} \phi\left(\frac{\mu_V}{\sqrt{\sigma_V^2 + 1}}\right).
\end{aligned}
\tag{B.3}$$

Equation (B.3) can be derived by straightforward derivation. The proof of Lemma 1 is thus completed by plugging (B.2) and (B.3) into (B.1).

B.2. Proof of Proposition 2

Recall that \bar{y}_i denotes the average outcome of patient i in all T treatment periods in an N-of-1 trial. Hence,

$$\begin{aligned}
E\left(\bar{y}_i\right) &= \frac{1}{T} \sum_{t=1}^T E\left(\alpha_i + \beta_i x_{it} + \epsilon_{it}\right) = \mu_A + \frac{1}{T} \sum_{t=1}^T E\left(\beta_i x_{it}\right) \\
&= \mu_A + \left(1 - \frac{m}{T}\right) E\left(\beta_i x_i^*\right).
\end{aligned}
\tag{B.4}$$

Equation (B.4) holds as because of balanced design $\sum_{i=1}^m x_{it} = 0$. Next, since $\hat{\beta}_i \sim N(\beta_i, \tau_i^2)$, we have

$$\begin{aligned}
E(\hat{\beta}_i x_i^*) &= E\left\{\beta_i E(x_i^* | \beta_i)\right\} = E\left\{\beta_i E\{2I(\hat{\beta}_i > 0) - 1 | \beta_i\}\right\} = E\left\{2\beta_i \Phi\left(\frac{\beta_i}{\tau_i}\right) - \beta_i\right\} \\
&= 2E\left\{\beta_i \Phi\left(\frac{\sqrt{m}\beta_i}{\sigma}\right)\right\} - \mu_B \\
&= 2\tau_i E\left\{\frac{\beta_i}{\tau_i} \Phi\left(\frac{\beta_i}{\tau_i}\right)\right\} - \mu_B \\
&= \mu_B \left[2\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - 1\right] + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right).
\end{aligned}$$

(B.5)

Expression (B.5) is obtained by applying Lemma 1 with $V = \beta_i/\tau_i$. Putting (B.5) into (B.4) gives

$$E(\bar{y}_i) = \mu_A + \left(1 - \frac{m}{T}\right) \left[\mu_B \left[2\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - 1\right] + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) \right]$$

(B.6)

thus completing the proof of Proposition 2.

B.3. Derivation of optimal experimentation length m^* and Main Result 1

For least squares $\hat{\beta}_i^{LS}$, the variance $\tau_i^2 = \lambda_i \sigma^2 / m$, where $\lambda_i = 1 + \sum_{s \neq i} x_{is} x_{is} / \rho_{si}$. Further supposing $\mu_B = 0$ simplifies (B.6) to

$$E(\bar{y}_i) = \mu_A + \left(1 - \frac{m}{T}\right) \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \lambda_i \sigma^2 / m}} \phi(0).$$

Hence, maximizing $E(\bar{y}_i)$ as a function of m is equivalent to maximizing the function

$$h(m) = \left(1 - \frac{m}{T}\right) \frac{1}{\sqrt{\xi_i + 1/m}}$$

where $\xi_i = \sigma_B^2 / (\lambda_i \sigma^2)$ is free of m . Using standard calculus arguments, we can show that the maximizer m^* of $h(m)$ solves the equation $2\xi_i m^{*2} + 3m^* - T = 0$ or equivalently,

$$m^* = \frac{\sqrt{9 + 8\xi_i T} - 3}{4\xi_i}.$$

(B.7)

The derivation of m^* is completed by multiplying $\sqrt{9 + 8\xi_i} + 3$ in the numerator and the denominator of (B.7), which gives

$$m^* = \frac{2T}{\sqrt{9 + 8\xi_i T} + 3}.$$

Now, since $\xi_i \geq 0$, we have $m^* \leq 2T/\sqrt{9} + 3 = T/3$. As a practical note, due to discreteness in m , the optimal m may be a result of rounding up m^* . Hence a slightly less sharp inequality would be $m^* \lesssim T/3 < T/3 + 1$.

Appendix C. Theoretical Results Concerning Power

In this section, we derive the expressions involved in the power of the Z -test—namely, Δ , $\text{var}(y_i^*)$, and $\text{var}(y_i')$.

Recall that p_0 and p_1 respectively denote the probabilities that the treating physicians will prescribe mexiletine ($x_{it} = -1$) and baclofen ($x_{it} = 1$) under the treatment program. Based on model (3.2), we can express the quality of an N-of-1 trial as:

$$E(y_i^*) = \frac{1}{T-m} \sum_{t=m+1}^T E(\alpha_t + \beta_t x_{it} + \epsilon_{it}) = \frac{1}{T-m} \sum_{t=m+1}^T E(\alpha_t + \beta_t x_i^* + \epsilon_{it}) = \mu_A + E(\beta_t x_i^*).$$

and analogously $E(y_i') = \mu_A + E(\beta_t x_i')$ where x_i' is the treatment given to patient i in SOC. Hence $\Delta = E(\beta_t x_i^*) - E(\beta_t x_i')$. Under the independence assumption of x_i' and β_t , we further obtain $E(y_i') = \mu_A + (2p_1 - 1)\mu_B$, and

$$\begin{aligned} \Delta &= E(\beta_t x_i^*) - \mu_B(2p_1 - 1) \\ &= \mu_B \left[2\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - 1 \right] + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - \mu_B(2p_1 - 1) \\ &= 2\mu_B \left[\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - p_1 \right] + \frac{2\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}} \phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) \end{aligned} \tag{C.1}$$

where $E(\beta_t x_i^*)$ is given in (B.5).

Next,

$$\begin{aligned} \text{var}(y_i^*) &= \text{var}\left(\alpha_t + \beta_t x_i^* + \sum_{t=m+1}^T \epsilon_{it} / (T-m)\right) \\ &= \sigma_A^2 + \text{var}(\beta_t x_i^*) + \frac{\sigma^2}{T-m} = \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \left\{ E(\beta_t x_i^*) \right\}^2 + \frac{\sigma^2}{T-m} \\ &= \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \left\{ \Delta + \mu_B(2p_1 - 1) \right\}^2 + \frac{\sigma^2}{T-m} \therefore = \sigma^{*2} \end{aligned}$$

The last equality is a result of (C.1). Similarly, we can show

$$\begin{aligned} \text{var}(y_i') &= \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \mu_B^2 (E x_i')^2 + \frac{\sigma^2}{T-m} \\ &= \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \mu_B^2 (2p_1 - 1)^2 + \frac{\sigma^2}{T-m}. \end{aligned}$$

Finally, under the null $\mu_B = 0$, we have

$$\frac{\Delta}{\sqrt{\text{var}(y_i^*) + \text{var}(y_i')}} = \frac{2\sigma_B^2\phi(0)}{\sqrt{\sigma_B^2 + \tau_i^2} \sqrt{2\sigma_A^2 + 2\sigma_B^2 - \frac{4\sigma_B^4\phi^2(0)}{\sigma_B^2 + \tau_i^2} + 2\sigma^2(T-m)}}.$$

Main Result 2 is proved by dividing σ_B^2 on the numerator and the denominator in the above expression, as a result of which the numerator will be a constant and the denominator will be a decreasing function σ_B^2 .

For the situations where the physicians have patient-specific knowledge to inform treatments under the SOC, we may postulate that

$$x_i' = \begin{cases} 2I(\beta_i > 0) - 1 & \text{with probability } \theta_c \\ 1 & \text{with probability } (1 - \theta_c)p_1 \\ -1 & \text{with probability } (1 - \theta_c)(1 - p_1). \end{cases} \quad (\text{C.2})$$

The parameter θ_c indicates how perfect the knowledge the physicians have about the specific best treatments for their patients, with $\theta_c = 1$ indicating perfect knowledge and $\theta_c = 0$ indicating no additional knowledge beyond the population-level information p_1 . Under the SOC treatment system (C.2), we have

$$E(\beta_i x_i') = E\{\beta_i E(x_i' | \beta_i)\} = 2\theta_c E\{\beta_i I(\beta_i > 0)\} - \theta_c \mu_B + (1 - \theta_c)\mu_B(2p_1 - 1) \quad (\text{C.3})$$

where

$$E\{\beta_i I(\beta_i > 0)\} = \sigma_B \phi(\mu_B/\sigma_B) + \mu_B \Phi(\mu_B/\sigma_B). \quad (\text{C.4})$$

Using (B.5), (C.3), and (C.4), after some algebra, we have

$$\begin{aligned} \Delta &= E\left(\beta_i x_i^*\right) - E\left(\beta_i x_i'\right) \\ &= 2\left(1 - \theta_c\right)\mu_B\left\{\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - p_1\right\} + \frac{2(1 - \theta_c)\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}}\phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) + \\ &\quad 2\theta_c\mu_B\left\{\Phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - \Phi\left(\frac{\mu_B}{\sigma_B}\right)\right\} + 2\theta_c\left[\frac{\sigma_B^2}{\sqrt{\sigma_B^2 + \tau_i^2}}\phi\left(\frac{\mu_B}{\sqrt{\sigma_B^2 + \tau_i^2}}\right) - \sigma_B\phi\left(\frac{\mu_B}{\sigma_B}\right)\right] \end{aligned}$$

It is instructive to consider the null case $\mu_B = 0$, under which

$$\Delta = \frac{2\sigma_B^2\phi(0)}{\sqrt{\sigma_B^2 + \tau_i^2}} - 2\theta_c\sigma_B\phi(0).$$

Hence, quality improvement Δ due to N-of-1 trials diminishes as the standard of care (SOC) patient-specific knowledge θ_c increases. Also under this special case, $\Delta > 0$ if and only if $\theta_c^2 < \sigma_B^2/(\sigma_B^2 + \tau_i^2)$

Similarly, we can obtain $\text{var}(y_i^*)$ and $\text{var}(y_i')$ under the SOC treatment system (C.2) by plugging (B.5) and (C.3) respectively in the followings:

$$\text{var}(y_i^*) = \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \{E(\beta_i x_i^*)\}^2 + \frac{\sigma^2}{T-m}$$

and

$$\text{var}(y_i') = \sigma_A^2 + \sigma_B^2 + \mu_B^2 - \{E(\beta_i x_i')\}^2 + \frac{\sigma^2}{T-m}$$

Then under the null $\mu_B = 0$, we have

$$\frac{\Delta}{\sqrt{\text{var}(y_i^*) + \text{var}(y_i')}} = \frac{2\sigma_B^2\phi(0) - 2\theta_c\sigma_B\sqrt{\sigma_B^2 + \tau_i^2}\phi(0)}{\sqrt{\sigma_B^2 + \tau_i^2} \sqrt{2\sigma_A^2 + 2\sigma_B^2 - \frac{4\sigma_B^4\phi^2(0)}{\sigma_B^2 + \tau_i^2} - 4\theta_c^2\sigma_B^2\phi^2(0)} + 2\sigma^2(T-m)}$$

By dividing σ_B^2 on the numerator and the denominator in the above expression, we can see that the numerator is an increasing function in σ_B^2 and the denominator is a decreasing function in σ_B^2 . Hence, Main Result 2 holds under this general case.

References

- Al-Chalabi A, & Hardiman O. (2013). The epidemiology of ALS: A conspiracy of genes, environment and time. *Nature Reviews Neurology*, 9(11), 617–628. 10.1038/nrneurol.2013.203 [PubMed: 24126629]
- Baldinger R, Katzberg HD, & Weber M. (2012). Treatment for cramps in amyotrophic lateral sclerosis/motor neuron disease. *Cochrane Database of Systematic Reviews*, Article CD004157. 10.1002/14651858.CD004157.pub2
- Caress JB, Ciarlone SL, Sullivan EA, Griffin LP, & Cartwright MS (2016). Natural history of muscle cramps in amyotrophic lateral sclerosis. *Muscle & Nerve*, 53(4), 513–517. 10.1002/mus.24892 [PubMed: 26332705]
- Cheung K, Wood D, Zhang K, Ridenour TA, Derby L, St Onge T, Duan N, Duer-Hefele J, Davidson KW, Kronish IM, & Moise N. (2020). Personal preferences for personalized trials among patients with chronic experience: An empirical Bayesian analysis of a conjoint survey. *BMJ Open*, 10(6), Article e036056. 10.1136/bmjopen-2019-036056
- Davidson KW, Silverstein M, Cheung K, Paluch RA, & Epstein LH (2021). Experimental designs to optimize treatments for individuals: Personalized N-of-1 trials. *JAMA Pediatrics*, 175(4), 404–409. 10.1001/jamapediatrics.2020.5801 [PubMed: 33587109]
- Edaravone [MCI-186] ALS 19 Study Group. (2017). Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: A randomised, double-blind, placebo-controlled trial. *Lancet Neurology*, 16(7), 505–512. 10.1016/s1474-4422(17)30115-1 [PubMed: 28522181]

- Hogben L, & Sim M. (1953). The self-controlled and self-recorded clinical trial for low-grade morbidity. *British Journal of Preventive and Social Medicine*, 7(4), 163–179. 10.1136/jech.7.4.163 [PubMed: 13106260]
- Kravitz RL, Duan N. (Eds), and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, & Vohra S) (2014). Design and implementation of N-of-1 trials: A user's guide. Agency for Healthcare Research and Quality. <https://effectivehealthcare.ahrq.gov/products/n-1-trials/research-2014-5>
- Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, Duan N, Wang Y, MacDonald S, Jerant A, Servadio JL, Haddad D, & Sim I. (2018). Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: A randomized clinical trial. *JAMA Internal Medicine*, 178(10), 1368–1378. 10.1001/jamainternmed.2018.3981 [PubMed: 30193253]
- Kronish IM, Cheung YK, Shimbo D, Julian J, Gallagher B, Parsons F, & Davidson KW (2019). Increasing the precision of hypertension treatment through personalized trials: A pilot study. *Journal of General Internal Medicine*, 34(6), 839–845. 10.1007/s11606-019-04831-z [PubMed: 30859504]
- Mitsumoto H, Brooks BR, & Silani V. (2014). Clinical trials in amyotrophic lateral sclerosis: Why so many negative trials and how can trials be improved? *Lancet Neurology*, 13(11), 1127–1138. 10.1016/s1474-4422(14)70129-2 [PubMed: 25316019]
- Mitsumoto H, Chiuhan C, Gilmore M, Zhang Y, Ibagon C, McHale B, Hupf J, & Oskarsson B. (2019). A novel muscle cramp scale (MCS) in amyotrophic lateral sclerosis (ALS). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 20(5–6), 328–335. 10.1080/21678421.2019.1603310 [PubMed: 31007059]
- Müller AR, Brands MMMG, van de Ven PM, Roes KCB, Cornel MC, van Karnebeek CDM, Wijburg FA, Daams JG, Boot E, & van Eeghen AM (2021). The power of 1: Systematic review of N-of-1 studies in rare genetic neurodevelopmental disorders. *Neurology*, 96(11), 529–540. 10.1212/WNL.00000000000011597 [PubMed: 33504638]
- Roustit M, Gai J, Gaget O, Khouri C, Mouhib M, Lotito A, Blaise S, Seinturier C, Subtil F, Paris A, Cracowski C, Imbert B, Carpentier P, Vohra S, & Cracowski J-L (2018). On-demand sildenafil as a treatment for raynaud phenomenon: A series of N-of-1 trials. *Annals of Internal Medicine*, 169(10), 694–703. 10.7326/m18-0517 [PubMed: 30383134]
- Samuel JP, Tyson JE, Green C, Bell CS, Pedroza C, Molony D, & Samuels J. (2019). Treating hypertension in children with n-of-1 trials. *Pediatrics*, 143(4), Article e20181818. 10.1542/peds.2018-1818
- Stunnenberg B, Raaphorst J, Groenewoud H, Statland J, Griggs R, Woertman W, Stegeman D, Timmermans J, Trivedi J, Matthews E, Saris C, Schouwenberg B, Drost G, van Engelen B, & van der Wilt G. (2018). A series of aggregated randomized-controlled N-of-1 trials with mexiletine in non-dystrophic myotonia: Clinical trial results and validation of rare disease design (p3.440) [70th Annual Meeting of the American-Academy-of-Neurology (AAN) ; Conference date: 21–04-2018 through 27–04-2018]. *Neurology*, 90(15 Suppl). https://n.neurology.org/content/90/15_Supplement/P3.440
- U.S. Food & Drug Administration. (2019). Adaptive designs for clinical trials of drugs and biologics: Guidance for Industry. <https://www.fda.gov/media/78495/download>
- van den Berg LH, Sorenson E, Gronseth G, Macklin EA, Andrews J, Baloh RH, Benatar M, Berry JD, Chio A, Corcia P, Genge A, Gubitza AK, Lomen-Hoerth C, McDermott CJ, Pioro EP, Rosenfeld J, Silani V, Turner MR, Weber M, . . . Mitsumoto H. (2019). Revised Airlie House consensus guidelines for design and implementation of ALS clinical trials. *Neurology*, 92(14), e1610–e1623. 10.1212/wnl.0000000000007242 [PubMed: 30850440]

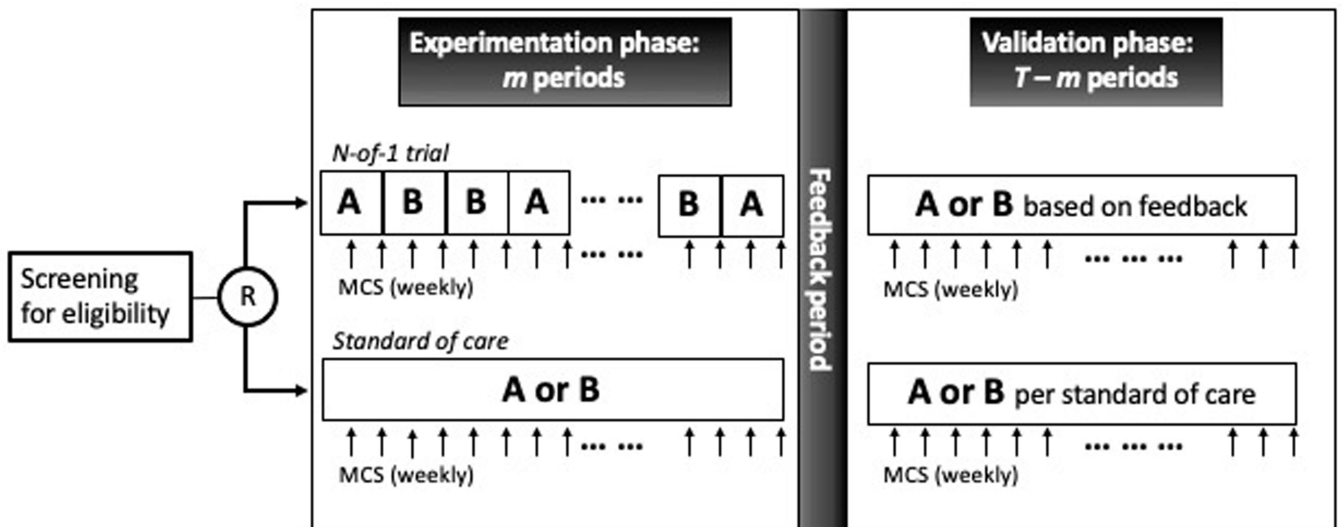


Figure 1. Schema of an evaluation program for personalized (N-of-1) trials comparing treatment A and treatment B. Under the evaluation program, patients are randomized to either an N-of-1 trial or the standard of care.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

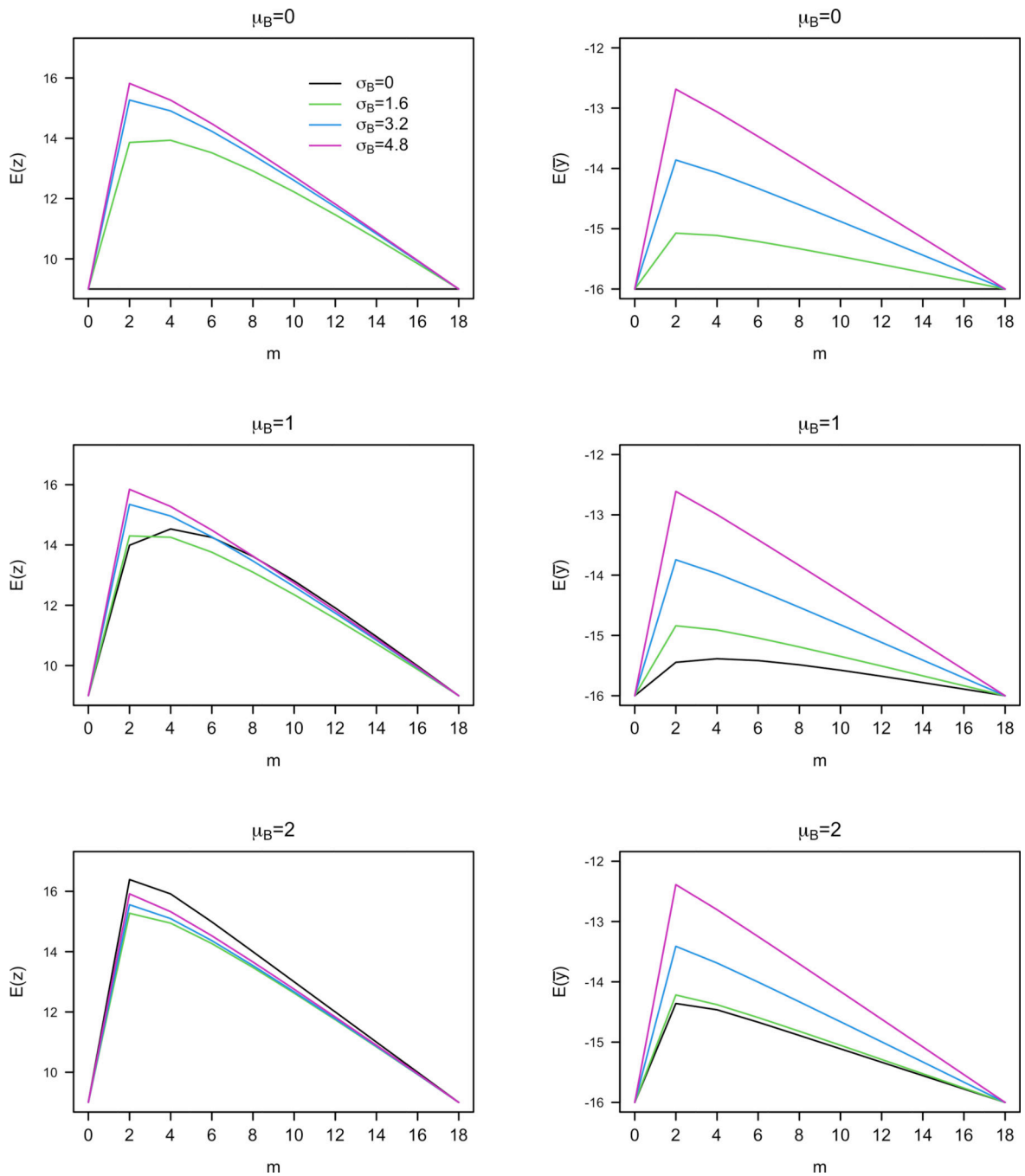


Figure 2. Patient-centric criteria vs experimentation length m under different values of μ_B and σ_B . Left: Expected number of optimal treatment periods vs m . Right: Expected average outcome (negative of MCS) of patient vs m .

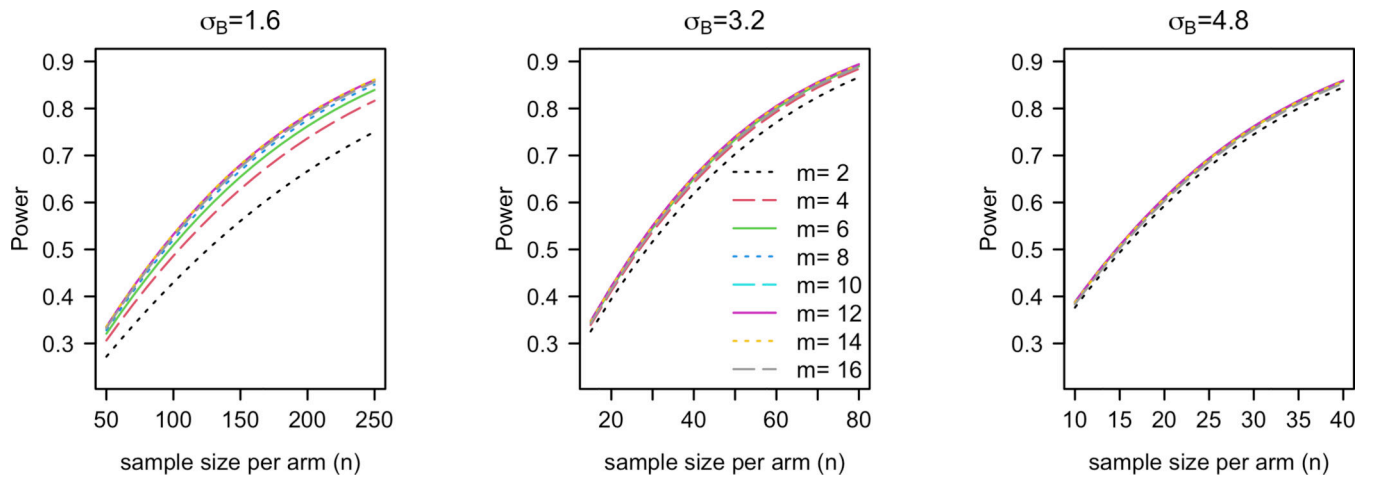


Figure 3. Power vs (n, m) for different values of σ_B with $\mu_B = 0$, $\sigma_A = 4.8$, $\sigma = 1.6$, $\rho = 0$, and $T = 18$.

Table 1.

Quality improvement Δ and power for comparing to a fully informed SOC (standard of care) with $p_1 = \Phi(\mu_B/\sigma_B)$ with $n = 34$, $m = 4$, $T = 18$, $\sigma_A = 4.8$, $\sigma_B = 4.8$, $\sigma = 1.6$ and $\rho = 0$.

μ_B	$p_1 = \Phi(\mu_B/\sigma_B)$	Δ	Power
0	0.50	3.8	80%
1.2	0.60	3.7	77%
1.6	0.63	3.6	75%
2.4	0.69	3.3	68%
4.8	0.84	2.3	39%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript