HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2025 July 01.

Published in final edited form as:

Biometrics. 2025 April 02; 81(2): . doi:10.1093/biomtc/ujaf061.

Learn-As-you-GO (LAGO) trials: optimizing treatments and preventing trial failure through ongoing learning

Ante Bing[1,*], **Donna Spiegelman**[2], **Daniel Nevo**[3], **Judith J. Lok**[1]

[1]Department of Mathematics and Statistics, Boston University, Boston, MA 02215, United States

[2]Department of Biostatistics, Yale University, New Haven, CT 06520, United States

[3]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, 6997801, Israel

Abstract

It is well known that changing the intervention package while a trial is ongoing does not lead to valid inference using standard statistical methods. However, it is often necessary to adapt, tailor, or tweak a complex intervention package in public health implementation trials, especially when the intervention package does not have the desired effect. This article presents conditions under which the resulting analyses remain valid even when the intervention package is adapted while a trial is ongoing. Our results on such Learn-As-you-GO (LAGO) trials extend the theory of LAGO for binary outcomes following a logistic regression model to LAGO for continuous outcomes under flexible conditional mean models. Because the mathematical methods for binary outcomes do not apply to continuous outcomes, the theory presented in this paper is entirely new. We derive point and interval estimators of the intervention effects and ensure the validity of hypothesis tests for an overall intervention effect. We develop a confidence set for the optimal intervention package, which achieves a pre-specified mean outcome while minimizing cost, and confidence bands for the mean outcome under all intervention package compositions. This work will be useful for the design and analysis of large-scale intervention trials where the intervention package is adapted, tailored, or tweaked while the trial is ongoing.

Keywords

adaptive clinical trial; dependent sample; implementation trial; large-scale intervention trial; public health

1 INTRODUCTION

Traditionally regulators have not allowed adaptive trial designs to be used for evidence in the drug approval process, but currently, the Food and Drug Administration (FDA) has put forward guidelines for adaptive clinical trials (FDA, 2016; 2019) and encourages trialists

[*]Corresponding author: Ante Bing, Department of Mathematics and Statistics, Boston University, 665 Commonwealth Ave, Boston, MA 02215, United States (abing@bu.edu).

CONFLICT OF INTEREST

None declared.

interested in adaptive trials to contact the FDA in the planning stage (FDA, 2019). While current adaptive clinical trial designs allow for changing randomization probabilities and dropping treatment arms (FDA, 2019), they do not allow for changes to an intervention package composition based on outcomes collected during the earlier stages of a trial. This limitation may have contributed to cases where large-scale intervention trials have "failed" (Stensland et al., 2014; Semrau et al., 2017; Fogel, 2018).

Learn-As-you-GO (LAGO) trials consist of $K > 1$ stages. The intervention package in a LAGO trial typically consists of multiple components. At the end of each stage, the data collected so far are combined and analyzed, and a systematically revised version of the intervention package is used in the next stage. Thus, as data are collected, the intervention package for later stages is systematically "learned" based on outcomes from previous stages. In the final analysis, the combined data from all stages are analyzed (see Section 3), and the test of no overall intervention effect is based on the data from all stages (see Section 5).

Nevo et al. (2021) provided the methodology for LAGO trials with binary outcomes, assuming a logistic regression model holds for the probability of success given the intervention package components, for which they prove consistency and asymptotic normality. However, their approach is limited to binary outcomes and logistic regression, necessitating a different theoretical framework for continuous outcomes and more flexible model specifications.

This article extends LAGO to continuous outcomes. To achieve this, we adopt a generalized linear model (GLM) framework, which refers to any semi-parametric model for how the conditional mean of an outcome depends on covariates and is linear in a link function, without additional restrictions on the outcome distribution. By extending LAGO to accommodate continuous outcomes under flexible conditional mean models, this article significantly broadens the application of LAGO across diverse implementation trials where outcomes are frequently measured on continuous scales, such as the percentage of clinical guidelines followed, patient satisfaction scores, or healthcare quality indicators.

When analyzing a LAGO trial, we cannot condition on the later-stage interventions, as this would imply conditioning on a function of the outcomes from earlier stages. Extending LAGO to continuous outcomes is not straightforward, as Nevo et al. (2021) used a coupling argument to prove the asymptotic properties of their estimators, and coupling cannot be generalized to continuous outcomes. To overcome this challenge, we assume throughout this article that the errors in the GLM (McCullagh and Nelder, 2019) are independent of the composition of the intervention package. Section 4 provides a detailed description of how we use an error-replacing strategy and empirical process theory to develop a new theoretical framework for LAGO with continuous outcomes.

The aims of LAGO trials are: (1) to estimate the impact of individual intervention package components on the outcome mean, (2) to test for an overall intervention effect when the intervention package has been "learned," and (3) to estimate the optimal intervention composition, $\mathbf{x}^{opt}$, achieving a pre-specified outcome mean while minimizing cost.

The multiphase optimization strategy (MOST) (Collins et al., 2007) and the sequential multiple assignment randomized trials (SMART) (Collins et al., 2007) are often compared to LAGO. The main differences are as follows. In MOST, only data from the last stage are used for testing and inference. SMART is a non-adaptive design, and each patient only learns from his/her own covariate and response history to estimate the optimal sequence of treatments. Web Appendix A provides more detailed comparisons.

LAGO has been identified as a promising approach in the field of implementation science (Beidas et al., 2022). These authors noted that LAGO can improve the alignment between implementation strategies and partner needs and contexts. Additionally, the user-centered design and approach of LAGO enables the optimization of implementation strategies and calibration of implementation support based on demonstrated need.

Section 8 retrospectively applies the LAGO design to the BetterBirth study (Hirschhorn et al., 2015; Semrau et al., 2017). The BetterBirth study was a costly failed trial, led by Harvard's Atul Gawande, which aimed to improve maternal and child health around the time of birth. By mimicking applying LAGO, we estimate the optimal intervention package that increases the percentage of essential birth practices (EBPs) performed to a pre-specified target outcome goal while minimizing cost. Our analysis differs from Nevo et al. (2021) in that they focused on a single binary outcome, the administration of oxytocin after delivery, whereas we considered the percentage of EBPs performed as a continuous outcome.

This article is organized as follows. Section 2 details the setting and notation. Section 3 describes the estimating equations for the proposed estimator for a LAGO trial with a GLM. Section 4 outlines the proofs of the asymptotic properties of the estimator. Section 5 covers testing in a LAGO trial. Section 6 describes the confidence sets and confidence bands for the optimal intervention. Section 7 describes simulations. Section 8 describes the LAGO analysis of the BetterBirth study. Section 9 discusses findings and future research topics.

2 SETTING, NOTATION, AND ASSUMPTIONS

Let Y_{ij} be the continuous outcome for patient i in center j. The optimal intervention is defined as the intervention that results in the mean outcome reaching a pre-specified outcome goal θ, while minimizing cost. Suppose the multi-component intervention package $\mathbf{x}$ has P components. Let $C(\mathbf{x})$ be a known continuous function representing the total cost of implementing the intervention package $\mathbf{x}$ (typically monetary, eg, in dollars). The cost function $C(\mathbf{x})$ is determined by subject-matter considerations. One approach is to identify the unit costs of the intervention components and define a linear cost function, where $C(\mathbf{x})$ consists of a fixed cost plus the sum of the products of dosage and unit cost for each intervention component. Another approach is to use the unit costs at the implemented dosage to consider a polynomial function of varying degrees to characterize the relationship between dosage and cost over the desired feasible ranges of the intervention components. One example of this approach, which has been favored by health economists on our team, is the cubic cost function, because it allows for an initial economy of scale followed by increased costs when the component levels exceed a threshold (see Web Appendix F.5).

Let $\mathbf{z}_j$ be fixed center characteristics for center j, such as hospital district or birth volume, that may be related to the outcome of interest. Throughout this article, we assume that centers do not participate in more than 1 stage.

Assumption 1:

We assume that the expected outcome of an individual i in center j with characteristic $\mathbf{z}_j$ under recommended intervention package $\mathbf{X}_j = \mathbf{x}_j$ and actual intervention package $\mathbf{A}_j = \mathbf{a}_j$, $E(Y_{ij} \mid \mathbf{A}_j = \mathbf{a}_j, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta}, \mathbf{z}_j)$ only depends on the recommended intervention $\mathbf{X}_j$ through the actual intervention $\mathbf{A}_j$ and follows a GLM

$$g\{E(Y_{ij} \mid \mathbf{A}_j = \mathbf{a}_j, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta}, \mathbf{z}_j)\} = \beta_0 + \boldsymbol{\beta}_1^T \mathbf{a}_j + \boldsymbol{\beta}_2^T \mathbf{z}_j, \tag{1}$$

where $g()$ is a twice continuously differentiable link function. $\boldsymbol{\beta}^T = \left(\beta_0, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T\right)$ are unknown parameters to be estimated from the data. Let $\boldsymbol{\beta}^*$ denote the true parameter vector of $\boldsymbol{\beta}$.

The optimal intervention package, which attains a pre-specified outcome goal while minimizing cost, for center j with baseline covariates $\mathbf{z}_j$ is the solution to the center-specific deterministic optimization problem

$$\mathrm{Min}_{\mathbf{x}_j} C(\mathbf{x}_j) \text{ subject to } \mathrm{E}(Y_{ij} \mid \mathbf{x}_j; \boldsymbol{\beta}, \mathbf{z}_j) \geq \theta, \tag{2}$$

with each component p of $\mathbf{x}_j$ in a pre-determined interval $[L_p, U_p]$, where $p = 1, \ldots, P$. Alternatively, instead of an absolute goal, the optimal intervention package could aim to increase the mean outcome for each center j by a pre-specified outcome goal $\Delta\theta$. In what follows, we focus on the case described in equation (2). In many settings, the intervention package implemented will not be center-specific. In such cases, the potential approaches are: (1) aiming for the average outcome over all centers to be at least θ, potentially optimized over a weighted average of center characteristics, and (2) aiming for all centers to reach the pre-specified outcome goal θ, thus implementing a single intervention package that enables all centers to achieve this goal. We have found this approach to be attractive to our current collaborators thus far. The second approach relates conceptually to maximin projection learning (Shi et al., 2018), which seeks a single treatment decision rule that performs adequately across all patient subgroups, even under the worst-case scenarios among different subgroups.

Calculating the estimated optimal intervention packages in LAGO trials (see Remarks 1 and 3) requires solving optimization problems defined by equation (2). These optimization problems substitute the estimates, $\hat{\boldsymbol{\beta}}$, for $\boldsymbol{\beta}$, consider $\mathbf{z}_j$ as fixed, so that the solutions are functions of $\hat{\boldsymbol{\beta}}$ and $\mathbf{z}_j$. We assume that there is a unique solution to equation (2).

For simplicity, we present the theory for LAGO trials with $K = 2$ stages. Web Appendix E extends the theory to $K > 2$ stages. In each stage $k, k = 1$ or 2, $n_j^{(k)}$ patients in center

$j, j = 1, \ldots, J^{(k)}$ are enrolled, with $J^{(k)}$ fixed. Let $n^{(k)} = \sum_{j=1}^{J^{(k)}} n_j^{(k)}$ be the number of patients in stage k, $n = n^{(1)} + n^{(2)}$ the total number of patients across the 2 stages, and $\alpha_{jk} = \lim_{n\to\infty} n_j^{(k)}/n > 0$, which assumes that the ratio between the sample size for center j in stage k and the total sample size converges to a non-zero constant as n goes to infinity. This assumption is a reasonable approximation if the number of patients in each center j of stages 1 and 2 is large, as in most large-scale implementation trials. At stage 1, the initial intervention package for each center, $\mathbf{x}_j^{(1)}$, is recommended based on the trial investigators' best guess, and/or based on pilot data before the trial starts. In stage 2, the recommended intervention package $\mathbf{X}_j^{(2,n^{(1)})}$ for center j is calculated from the stage 1 data and other accumulated knowledge, possibly including summary measures of, for example, qualitative information from providers, patients, and other stakeholders. The superscript $(2, n^{(1)})$ indicates that $\mathbf{X}_j^{(2,n^{(1)})}$ depends on the data from the $n^{(1)}$ patients in stage 1. Often, $\mathbf{X}_j^{(2,n^{(1)})}$ will solve equation (2) with the estimator $\hat{\boldsymbol{\beta}}^{(1)}$ in place of $\boldsymbol{\beta}$ (see Remark 1). Let $\mathbf{A}_j^{(2,n^{(1)})}$ be the actual intervention package implemented in center j in stage 2. Under perfect adherence, $\mathbf{A}_j^{(2,n^{(1)})} = \mathbf{X}_j^{(2,n^{(1)})}$. In practice, both stage 1 and stage 2 centers may not adhere completely to all the intervention components. We assume that to the extent it is not random, adherence depends on the recommendation $\mathbf{X}_j^{(2,n^{(1)})}$ and on the stage 2 center characteristics $\mathbf{z}_j^{(2)}$, but not on other predictors of the outcome.

Remark 1:

One way to calculate the stage 2 recommended interventions $\mathbf{X}_j^{(2,n^{(1)})}$, *denoted by* $\hat{\mathbf{x}}_j^{opt,(2,n^{(1)})}$, *is by using the following function* f. *This function* f *takes the stage 1-based estimate* $\hat{\boldsymbol{\beta}}^{(1)}$, *the pre-specified goal* θ, *and the pre-specified stage 2 center characteristics* $\mathbf{z}_j^{(2)}$ *as input, solves the optimization problem of* equation (2) *by using* $\hat{\boldsymbol{\beta}}^{(1)}$ *in place of* $\boldsymbol{\beta}$, $\mathbf{z}_j^{(2)}$ *in place of* $\mathbf{z}_j$, *and returns center-specific recommended interventions as output. We denote* $\hat{\boldsymbol{x}}_j^{opt,(2,n^{(1)})} = f(\hat{\boldsymbol{\beta}}^{(1)}; \mathbf{z}_j^{(2)})$, *with* f *solving* equation (2).

The optimization algorithm used to solve equation (2) *varies depending on the form of the cost function. Details of the optimization algorithms for solving* equation (2) *are provided in* Web Appendix B.

For clarity of the expositions, below we will work with $\hat{\mathbf{x}}_j^{opt,(2,n^{(1)})}$ as the recommended intervention. Let $Y_{ij}^{(1)}$ be the outcome of participant i in center j of stage 1. Since stage 1 centers may not adhere completely to all the intervention components, let $\mathbf{a}_j^{(1)} = h_j^{(1)}(\mathbf{x}_j^{(1)})$ be the actual intervention package for center j of stage 1, where $h_j^{(1)}$ is a continuous, deterministic function for each center j in stage 1. In practice, both $\mathbf{a}_j^{(1)}$ and $\mathbf{x}_j^{(1)}$ are observed, but the $h_j^{(1)}$ are typically unknown. Let $\mathbf{Y}_j^{(1)} = \left(Y_{1j}^{(1)}, \ldots, Y_{n_j^{(1)}j}^{(1)}\right)$ be the outcomes of patients $1, \ldots, n_j^{(1)}$ in center j of stage 1. Let $\overline{\mathbf{a}}^{(1)} = \left(\mathbf{a}_1^{(1)}, \ldots, \mathbf{a}_{J^{(1)}}^{(1)}\right)$, $\overline{\mathbf{z}}^{(1)} = \left(\mathbf{z}_1^{(1)}, \ldots, \mathbf{z}_{J^{(1)}}^{(1)}\right)$, and $\overline{\mathbf{Y}}^{(1)} = \left(\mathbf{Y}_1^{(1)}, \ldots, \mathbf{Y}_{J^{(1)}}^{(1)}\right)$ be the actual interventions, center characteristics, and outcomes for center $1, \ldots, J^{(1)}$ of stage 1,

respectively. Let $\overline{\hat{\mathbf{x}}}^{opt,\left(2,n^{(1)}\right)}=\left(\hat{\mathbf{x}}_1^{opt,\left(2,n^{(1)}\right)},\ldots,\hat{\mathbf{x}}_{J^{(2)}}^{opt,\left(2,n^{(1)}\right)}\right)$ be the recommended interventions for center $1,\ldots,J^{(2)}$ of stage 2.

Let $Y_{ij}^{\left(2,n^{(1)}\right)}, i=1,\ldots,n_j^{(2)}$ be the outcome of participant i in center j of stage 2. Let $\mathbf{Y}_j^{\left(2,n^{(1)}\right)}=\left(Y_{1j}^{\left(2,n^{(1)}\right)},\ldots,Y_{n_j^{(2)}j}^{\left(2,n^{(1)}\right)}\right)$ be the outcomes of patients $1,\ldots,n_j^{(2)}$ in center j of stage 2. Since stage 2 centers may not adhere completely to all the intervention components, let $\mathbf{A}_j^{\left(2,n^{(1)}\right)}=h_j^{(2)}\left(\hat{\mathbf{x}}_j^{opt,\left(2,n^{(1)}\right)}\right)$ be the actual intervention package for center j of stage 2, where $h_j^{(2)}$ is a continuous, deterministic function for each center j in stage 2. In practice, both $\mathbf{A}_j^{\left(2,n^{(1)}\right)}$ and $\hat{\mathbf{x}}_j^{opt,\left(2,n^{(1)}\right)}$ are observed, but the $h_j^{(2)}$ are typically unknown. Let $\overline{\mathbf{A}}^{\left(2,n^{(1)}\right)}=\left(\mathbf{A}_1^{\left(2,n^{(1)}\right)},\ldots,\mathbf{A}_{J^{(2)}}^{\left(2,n^{(1)}\right)}\right)$, $\overline{\mathbf{z}}^{(2)}=\left(\mathbf{z}_1^{(2)},\ldots,\mathbf{z}_{J^{(2)}}^{(2)}\right)$, and $\overline{\mathbf{Y}}^{\left(2,n^{(1)}\right)}=\left(\mathbf{Y}_1^{\left(2,n^{(1)}\right)},\ldots,\mathbf{Y}_{J^{(2)}}^{\left(2,n^{(1)}\right)}\right)$ be the actual interventions, center characteristics, and outcomes for each center $1,\ldots,J^{(2)}$ of stage 2, respectively.

Assumption 2:

Conditionally on $\overline{\hat{\mathbf{x}}}^{opt,\left(2,n^{(1)}\right)}$, $\left(\overline{\mathbf{A}}^{\left(2,n^{(1)}\right)},\overline{\mathbf{Y}}^{\left(2,n^{(1)}\right)}\right)$ *is independent of the stage 1 data* $\left(\overline{\mathbf{a}}^{(1)},\overline{\mathbf{Y}}^{(1)}\right)$. *That is, learning from data from earlier stages is only through the determination of the recommended intervention*.

Assumption 3:

For each center $j=1,\ldots,J^{(2)}$, *the stage 2 recommended intervention* $\hat{\boldsymbol{x}}_j^{opt,\left(2,n^{(1)}\right)}$ *converges in probability to a center-specific limit* $\mathbf{x}_j^{(2)}$.

Remark 2:

Assumption 3 *holds, for example, when in an open neighborhood of the true* $\boldsymbol{\beta}^*$, *the recommended intervention is a continuous function of either (1) the maximum likelihood estimator (MLE) of* $\boldsymbol{\beta}$ *based on the stage 1 data; (2) a solution of the generalized estimating equation [GEE,* Liang and Zeger (1986)] *of* $\boldsymbol{\beta}$ *based on the stage 1 data; or (3) averages based on all stage 1 patients. Thus*, Assumption 3 *holds if* $\hat{\boldsymbol{x}}_j^{opt,\left(2,n^{(1)}\right)}$ *solves* equation (2) *and such solution is unique for the true* $\boldsymbol{\beta}^*$. *If no solution* $\mathbf{x}$ *exists that achieves the outcome goal* θ *from* equation (2), Assumption 3 *still holds when using the method described in* Section 5.1 *of the* Supplementary Material *of* Nevo et al. (2021) *to calculate* $\hat{\boldsymbol{x}}_j^{opt,\left(2,n^{(1)}\right)}$. *The intuition behind this method is to shrink the recommended intervention toward the stage 1 recommended intervention using a continuous function of the stage 1-based estimate* $\hat{\boldsymbol{\beta}}^{(1)}$.

Under Assumption 3, the assumption on the h_j that map recommended interventions to actual interventions in center j, and the Continuous Mapping Theorem, it follows that $\mathbf{A}_j^{\left(2,n^{(1)}\right)}=h_j^{(2)}\left(\hat{\mathbf{x}}_j^{opt,\left(2,n^{(1)}\right)}\right)$ converges in probability to $\mathbf{a}_j^{(2)}=h_j^{(2)}\left(\mathbf{x}_j^{(2)}\right)$.

Assumption 4:

The covariates $\mathbf{z}_j$*, the outcomes* Y_{ij}*, and the parameter space for* $\boldsymbol{\beta}$ *all take values in a compact space.*

Assumption 5:

Let ϵ_{ij} *be the error term for individual* i *in center* j *of the GLM from* Assumption 1. *That is,* $Y_{ij} = g^{-1}(\beta_0 + \boldsymbol{\beta}_1^T \mathbf{a}_j + \boldsymbol{\beta}_2^T \mathbf{z}_j) + \epsilon_{ij}$*, with* $E(\epsilon_{ij} \mid \mathbf{a}_j; \mathbf{z}_j) = 0$*. The errors* ϵ_{ij} *are independent, and their distributions are independent of the intervention package composition* $\mathbf{a}_j$*, although they may depend on* $\mathbf{z}_j$*. We denote* $\sigma^2(\mathbf{z}_j) = VAR(\epsilon_{ij}; \mathbf{z}_j)$*. Often, the* ϵ_{ij} *will be assumed to be i.i.d., with distribution not dependent on the fixed* $\mathbf{z}_j$*.*

3 ESTIMATING β AND THE OPTIMAL INTERVENTION PACKAGE

This section describes the estimating equations for the proposed estimator $\hat{\boldsymbol{\beta}}$, and how to subsequently use $\hat{\boldsymbol{\beta}}$ to estimate the optimal intervention package, which attains a pre-specified outcome goal while minimizing cost. The adaption of the intervention package based on prior outcomes causes dependence between stages, invalidating standard statistical theory. The usual (non-LAGO) estimator $\hat{\boldsymbol{\beta}}$ for the true parameter $\boldsymbol{\beta}^*$ is the solution to the GEE under an independence working correlation structure (Liang and Zeger, 1986)

$$
\begin{aligned}
0 = \mathbf{U}^{(g)}(\boldsymbol{\beta}) = \frac{1}{n}\Bigg[& \sum_{j=1}^{J^{(1)}} \sum_{i=1}^{n_j^{(1)}} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}\left(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}\right) \right\} \\
& \left\{ Y_{ij}^{(1)} - g^{-1}\left(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}\right) \right\} \\
& + \sum_{j=1}^{J^{(2)}} \sum_{i=1}^{n_j^{(2)}} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^{-1}} g^{-1}\left(\mathbf{A}_j^{\left(2, n^{(1)}\right)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}\right) \right\} \\
& \left\{ Y_{ij}^{\left(2, n^{(1)}\right)} - g^{-1}\left(\mathbf{A}_j^{\left(2, n^{(1)}\right)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}\right) \right\} \Bigg].
\end{aligned}
\tag{3}
$$

We estimate $\boldsymbol{\beta}$ in a LAGO trial by the solution to equation (3). The superscript (g) in $\mathbf{U}^{(g)}(\boldsymbol{\beta})$ reminds us of that $\mathbf{U}^{(g)}(\boldsymbol{\beta})$ are estimating equations for the LAGO GLM with a general link function. Equation (3) only involves $\mathbf{A}_j^{\left(2, n^{(1)}\right)}$, not $\hat{\mathbf{x}}_j^{opt, \left(2, n^{(1)}\right)}$, because under Assumption 1, the expected outcome of individual i in center j with center characteristics $\mathbf{z}_j$ under recommended intervention package $\mathbf{X}_j$ and actual intervention package $\mathbf{A}_j$ only depends on the recommended intervention $\mathbf{X}_j$ through the actual intervention $\mathbf{A}_j$. Asymptotic theory for $\hat{\boldsymbol{\beta}}$ is complicated by the fact that the stage 2 interventions $\overline{\mathbf{A}}^{\left(2, n^{(1)}\right)}$ depend on the stage 1 outcomes $\overline{\mathbf{Y}}^{(1)}$, so the 2 terms in (3) are not independent.

Section 4 shows that despite these dependencies, under the assumptions of Section 2, the estimator $\hat{\boldsymbol{\beta}}$ solving equation (3) is both consistent and asymptotically normal.

Remark 3:

To estimate the final optimal intervention package for a given center with center characteristics $\tilde{\mathbf{z}}$, *we use the function* f *from* Remark 1. *This function now takes the estimate* $\hat{\boldsymbol{\beta}}$ *based on the combined data from stage 1 and stage 2 [solving* equation (3)], *the pre-specified outcome goal* θ, *and the pre-specified center characteristics* $\tilde{\mathbf{z}}$ *as input, solves the optimization problem of* equation (2) *by using* $\hat{\boldsymbol{\beta}}$ *in place of* $\boldsymbol{\beta}$, *and* $\tilde{\mathbf{z}}$ *in place of* $\mathbf{z}_j$, *and then returns an optimal intervention package. We denote* $\hat{\mathbf{x}}^{opt}(\tilde{\mathbf{z}}) = f(\hat{\boldsymbol{\beta}}; \tilde{\mathbf{z}})$, *with* f *solving* equation (2). *Note that* $\tilde{\mathbf{z}}$ *can take values observed in the current study, or new values not observed if extrapolation is possible*.

4 ASYMPTOTIC PROPERTIES OF $\hat{\beta}$

We derive the asymptotic properties of the final estimator $\hat{\boldsymbol{\beta}}$ as n, the total number of patients across the stages, goes to infinity. As in Section 2, we assume that the limiting proportion of patients in center j of stage k exists and is denoted by $\alpha_{jk} = \lim_{n\to\infty} n_j^{(k)}/n$. Consider a 2-stage LAGO design with fixed number of centers $J^{(k)}$ in stage k. Nevo et al. (2021) provide a proof of consistency and asymptotic normality for $\hat{\boldsymbol{\beta}}$ under logistic regression. In contrast, the methods developed here apply to a flexible conditional mean model, the link function $g()$ does not need to be the canonical link function, and most importantly, the coupling arguments previously used for binary outcomes to show consistency and asymptotic normality do not generalize to continuous outcomes. Thus, the proofs outlined in this section are fundamentally different from those in Nevo et al. (2021).

Theorem 1:

(Consistency). Under Assumptions 1–5, $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}^*$.

To prove consistency of $\hat{\boldsymbol{\beta}}$, we show that in spite of the fact that equation (3) does not consist of i.i.d. terms, Theorem 5.9 of van der Vaart (2000) can be used. We show that the 2 conditions of this theorem are satisfied. For the first condition of Theorem 5.9 of van der Vaart (2000), we show that

$$\sup_{\beta} \left\| \mathbf{U}^{(g)}(\boldsymbol{\beta}) - \mathbf{u}^{(g)}(\boldsymbol{\beta}) \right\| \xrightarrow{P} 0, \tag{4}$$

where $\mathbf{u}^{(g)}(\boldsymbol{\beta})$ are the expected values of the estimating equations under the limiting design, in particular,

$$\mathbf{u}^{(g)}(\boldsymbol{\beta}) = \sum_{j=1}^{J^{(1)}} \alpha_{j1} \left[\left\{ \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}\left(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}\right) \right\} \left\{ g^{-1}\left(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}^*, \mathbf{z}_j^{(1)}\right) - g^{-1}\left(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}\right) \right\} \right] + \sum_{j=1}^{J^{(2)}} \alpha_{j2} \left[\left\{ \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}\left(\mathbf{a}_j^{(2)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}\right) \right\} \left\{ g^{-1}\left(\mathbf{a}_j^{(2)}; \boldsymbol{\beta}^*, \mathbf{z}_j^{(2)}\right) - g^{-1}\left(\mathbf{a}_j^{(2)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}\right) \right\} \right], \tag{5}$$

where $\mathbf{A}_j^{\left(2, n^{(1)}\right)} \xrightarrow{P} \mathbf{a}_j^{(2)}$.

To prove equation (4), Web Appendix B shows that $\mathbf{U}^{(g)}(\boldsymbol{\beta}) - \mathbf{u}^{(g)}(\boldsymbol{\beta})$ can be decomposed into 5 components. Let $\epsilon_{ij}^{(2)}$ be the error that patient i in center j would experience under the limiting intervention $\mathbf{a}_j^{(2)}$. By Assumption 5, replacing the $\epsilon_{ij}^{\left(2, n^{(1)}\right)}$ with $\epsilon_{ij}^{(2)}$ does not change the distribution of the part in $\mathbf{U}^{(g)}(\boldsymbol{\beta}) - \mathbf{u}^{(g)}(\boldsymbol{\beta})$ that includes $\epsilon_{ij}^{\left(2, n^{(1)}\right)}$. Combining the error replacing approach with the concept of Donsker classes from empirical process theory, the supremum over $\boldsymbol{\beta}$ of each of the 5 components of $\mathbf{U}^{(g)}(\boldsymbol{\beta}) - \mathbf{u}^{(g)}(\boldsymbol{\beta})$ converges to 0 in probability. Then, by the triangle inequality, equation (4) holds.

For LAGO to work properly, we need variation in the intervention components to estimate the treatment effect parameters. The uniqueness of $\boldsymbol{\beta}^*$ as a maximizer or 0 of $\mathbf{u}^{(g)}(\boldsymbol{\beta})$ [equation (5)] has been studied by various authors, see, for example, Chapter 2.2 of Fahrmeir and Tutz (2013). $\mathbf{u}^{(g)}(\boldsymbol{\beta})$ is the same as if all interventions $\mathbf{a}_j^{(1)}$ and $\mathbf{a}_j^{(2)}$ were fixed before the trial; so, if there is enough variation in those intervention components, the second condition in Theorem 5.9 of van der Vaart (2000) is also satisfied and we conclude that $\hat{\boldsymbol{\beta}}$ is consistent.

Theorem 2:

(Asymptotic normality). Under Assumptions 1–5,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right) \xrightarrow{D} N\left(0, J(\boldsymbol{\beta}^*)^{-1} V(\boldsymbol{\beta}^*) J(\boldsymbol{\beta}^*)^{-1}\right), \tag{6}$$

where

$$J(\boldsymbol{\beta}^*) = \sum_{j=1}^{J^{(1)}} \alpha_{j1} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}^*} g^{-1}(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}) \right)^{\otimes 2} + \sum_{j=1}^{J^{(2)}} \alpha_{j2} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}^*} g^{-1}(\mathbf{a}_j^{(2)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}) \right)^{\otimes 2},$$

$$V(\boldsymbol{\beta}^*) = \sum_{j=1}^{J^{(1)}} \alpha_{j1} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}^*} g^{-1}(\mathbf{a}_j^{(1)}; \boldsymbol{\beta}, \mathbf{z}_j^{(1)}) \right)^{\otimes 2} \sigma^2(\mathbf{z}_j^{(1)}) + \sum_{j=1}^{J^{(2)}} \alpha_{j2} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}^*} g^{-1}(\mathbf{a}_j^{(2)}; \boldsymbol{\beta}, \mathbf{z}_j^{(2)}) \right)^{\otimes 2} \sigma^2(\mathbf{z}_j^{(2)}),$$

and $\otimes 2$ *is the Kronecker product.*

The variance in equation (6) can be estimated by replacing $\boldsymbol{\beta}^*$, $\mathbf{a}_j^{(2)}$, α_{j1}, and α_{j2} with $\hat{\boldsymbol{\beta}}$, $\mathbf{A}_j^{(2,n^{(1)})}$, $n_j^{(1)}/n$, and $n_j^{(2)}/n$, respectively.

The proof of Theorem 2 uses a combination of strategies and concepts: building on Theorem 1, applying the mean value theorem, employing the error-replacement strategy from the proof of Theorem 1, invoking Donsker classes from empirical process theory, and leveraging Lévy's Continuity Theorem for characteristic functions. Web Appendix C provides a detailed proof for Theorem 2. Web Appendix D presents the proofs for Theorem 1 and Theorem 2 using a log link function as a concrete example.

5 HYPOTHESIS TESTING

In a LAGO trial, one of the main objectives is to evaluate the null hypothesis of no overall intervention effect. This null hypothesis, represented by $H_0 : \boldsymbol{\beta}_1 = 0$, can be tested using standard methods, such as the P degrees-of-freedom χ^2 test. The validity of the χ^2 test for this purpose is guaranteed by Theorem 2. In LAGO trials, let R be the group indicator that identifies the intervention group ($R = 1$) and the control group ($R = 0$). Let μ_0 and μ_1 represent the mean outcome values in the control and intervention groups, respectively. An alternative test for $H_0: \boldsymbol{\beta}_1 = 0$ is to test the implied $\widetilde{H}_0: \mu_0 = \mu_1$. Under the null hypothesis, learning does not affect the distribution of the outcomes, because the stage 2 intervention has no effect on the stage 2 outcomes. Under the null hypothesis, the distribution of the outcomes is the same regardless of the intervention, and we can use any standard 1-degree-of-freedom test (such as the Z-test) to compare the distribution of the outcomes in the intervention and the control group. This leads to a test with the usual level of significance (either exact or asymptotic, depending on the test used).

One can also take center characteristics $\mathbf{z}$ into account when testing $H_0: \boldsymbol{\beta}_1 = 0$ in a LAGO trial, by testing whether $\gamma = 0$ in the model $g(E(Y \mid \gamma; \boldsymbol{\beta}, \mathbf{z})) = \beta_0 + \boldsymbol{\beta}_2^T \mathbf{z} + \gamma R$. Under the null hypothesis, as before, there is independence between the stages, and $\boldsymbol{\beta}_1 = 0$ implies that $\gamma = 0$.

6 CONFIDENCE SETS AND CONFIDENCE BANDS

The confidence set for the optimal intervention package is a list of intervention package compositions that can be expected to include the optimal intervention in 95% of such trials. To construct this confidence set, we create a confidence interval for $\mu = E(Y \mid \mathbf{x}, \tilde{\mathbf{z}}; \boldsymbol{\beta}^*)$ for

a given value $\tilde{\mathbf{z}}$ and for each possible value of $\mathbf{x}$. We first calculate a 95% confidence interval for $g(\mu)$ as $CI_{g(\mu)} = \left(1\mathbf{x}^T\tilde{\mathbf{z}}^T\right)\hat{\boldsymbol{\beta}} \pm 1.96\sqrt{\sigma^2_\mu(\hat{\boldsymbol{\beta}};\mathbf{x},\tilde{\mathbf{z}})}$ where $\sigma^2_\mu(\hat{\boldsymbol{\beta}};\mathbf{x},\tilde{\mathbf{z}})$ can be calculated based on Theorem 2: $\sigma^2_\mu(\hat{\boldsymbol{\beta}};\mathbf{x},\tilde{\mathbf{z}}) = \left(1\mathbf{x}^T\tilde{\mathbf{z}}^T\right)n^{-1}\hat{J}\left(\hat{\boldsymbol{\beta}}\right)^{-1}\hat{V}\left(\hat{\boldsymbol{\beta}}\right)\hat{J}\left(\hat{\boldsymbol{\beta}}\right)^{-1}\left(1\mathbf{x}^T\tilde{\mathbf{z}}^T\right)^T$. It follows that the 95% confidence interval for $E(Y \mid \mathbf{x};\boldsymbol{\beta}^*,\tilde{\mathbf{z}})$ is $CI_\mu = g^{-1}(CI_{g(\mu)})$. The confidence set can then be calculated as $CS\left(\mathbf{x}^{opt}\right) = \{\mathbf{x}: CI_\mu \ni \theta\}$. That is, $CS\left(\mathbf{x}^{opt}\right)$ includes all intervention packages $\mathbf{x}$ for which the outcome goal θ [equation (2)] is inside the confidence interval for the mean outcome under intervention package $\mathbf{x}$. By Theorem 2, the confidence set $CS\left(\mathbf{x}^{opt}\right)$ contains $\mathbf{x}^{opt}$ with asymptotic probability at least 0.95.

Next, we construct confidence bands for the mean outcome under all different intervention package compositions. These confidence bands have asymptotic 95% coverage simultaneously for all feasible intervention package compositions. To calculate the confidence bands for the expected outcomes $E(Y \mid \mathbf{x};\boldsymbol{\beta}^*,\tilde{\mathbf{z}})$, we first compute the 95% confidence bands for $\left(1\mathbf{x}^T\tilde{\mathbf{z}}^T\right)\boldsymbol{\beta}^*$, similar to Scheffe (1999) and Nevo et al. (2021) (Section 4 of their Supplementary Materials),

$$CB_{g(\mu)} = \left(1\mathbf{x}^T\tilde{\mathbf{z}}^T\right)\hat{\boldsymbol{\beta}} \pm \sqrt{\chi^2_{0.95,\,p+q+1}\sigma^2_\mu(\hat{\boldsymbol{\beta}};\mathbf{x},\tilde{\mathbf{z}})}, \tag{7}$$

where $\chi^2_{0.95,\,p+q+1}$ is the 95 th percentile of the χ^2_{p+q+1} distribution. $p+q+1$ is the dimension of $\boldsymbol{\beta}$, where p and q are the dimensions of $\mathbf{a}$ and $\mathbf{z}$, respectively. The confidence bands $CB_\mu = g^{-1}(CI_{g(\mu)})$ guarantee asymptotic simultaneous 95% coverage for the mean outcome under all intervention package compositions.

7 SIMULATIONS

Our methods were evaluated through simulation studies, which were organized into 3 parts, each based on 2000 simulated datasets. Simulation 1 considered 2 scenarios of 2-stage LAGO designs.

Simulation 1 scenario 1a had the same number of centers J in both the intervention and the control arm for both stage 1 and stage 2. We considered $J = 6,\ 10,\ 20$, with $n_j^{(1)} = 50,\ 100$, and $n_j^{(2)} = 100,\ 200$. The intervention consisted of 2 components, $\mathbf{x} = (x_1, x_2)$, with the minimum and maximum values of x_1 and x_2 being $[L_1, U_1] = [0,2]$, and $[L_2, U_2] = [0,\ 8]$, respectively. The recommended interventions for stage 1 were set to be the middle of the range for the 2 intervention components, $\mathbf{x}^{(1)} = \left(1,\ 4\right)$. The model for the continuous outcome was $g(E(Y_{ij} \mid \mathbf{a}_j;\boldsymbol{\beta},\mathbf{z}_j)) = \boldsymbol{\beta}_1^T\mathbf{a}_j + \boldsymbol{\beta}_2^T\mathbf{z}_j$, with $g()$ the logit link function. The logit link function was chosen because it was used in the analysis of the BetterBirth study of Section 8 to restrict the expected fractions to values between 0 and 1. The true coefficient values for the intervention components $\boldsymbol{\beta}_1^*$ were set as $(\beta_{11}^*, \beta_{12}^*)$ based on the estimated parameter values and their confidence intervals in the final analysis of the BetterBirth study (Table 2):(0.1863, 0.15), (0.0438, 0.17), (0.1, 0.2133), and (0.1062, 0.16). The exact true coefficient values were

chosen to facilitate the confidence set and confidence bands calculations. More specifically, ($\beta_{11}^* = 0.1863, \beta_{12}^* = 0.15$) was chosen so that the true optimal intervention was (1, 8) within 3 decimal places. Other β_{11}^* and β_{12}^* values were chosen for the same reason. A baseline center characteristic $Z \sim N(0,\ 1)$ was also included with a true coefficient value of $\beta_z^* = -0.2$. The outcome goal was $\theta = 0.8$, and the optimization problem was solved as described in equation (2) to obtain the recommended interventions. For simplicity, no intercept was included in the models.

In simulation 1 scenario 1b, instead of having the same number of centers in both stages, fewer centers and a smaller per-center sample size were included in stage 1 than in stage 2. Additional details for simulation 1 are presented in Web Appendix F.1.

Table 1a–c present selected results for simulation 1 scenarios 1a and 1b using a linear cost function with per unit cost for the two intervention components: $C = (c_1 = 8, c_2 = 2)$. Complete linear cost function results are presented in Web Appen dix F.2–F.4, while Web Appendix F. 5 provides selected results with a cubic cost function.

Table 1a shows that for $J > 6$, both $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ had minimal relative bias, and $\hat{\beta}_{12}$ had smaller relative bias than $\hat{\beta}_{11}$. While the ratio between the mean of the estimated standard error and the empirical standard error ranged from 0.7 to 0.9, the empirical coverage rates of the 95% confidence intervals for both β_{11} and β_{12} were close to 95%. From Table 1b, the bias and root mean squared error were small for both estimated optimal intervention components, with the final estimated optimal intervention based on all data (shown under “Stage 2/LAGO optimized”) having smaller bias and root mean squared error than the estimated optimal intervention based on stage 1 data (shown under “Stage 1”). Table 1c shows that the true mean under the final estimated optimal intervention was close to the outcome goal of 0.8 in most simulated datasets (see Table 1c column MeanOpt2). The coverage rate for both the confidence set for the optimal intervention and the simultaneous confidence bands for the intervention package components were very close to 95%.

Simulation 2 closely mimicked the BetterBirth study as if a LAGO design had been used. Details of simulation 2 are presented in Web Appendix F.6. Simulation 3 evaluated the power advantages of the LAGO design over the factorial design and MOST. Details and results are presented in Web Appendix F.7. The LAGO design had minimal finite sample bias and notably higher power than both the factorial design and MOST.

8 ILLUSTRATIVE EXAMPLE: THE BETTERBIRTH STUDY

We illustrated the LAGO design for continuous outcomes by retrospectively applying the LAGO design to the BetterBirth study (Hirschhorn et al., 2015; Semrau et al., 2017). The BetterBirth study was a costly failed trial, aiming to improve maternal and neonatal health outcomes in Uttar Pradesh, India by implementing the World Health Organization’s Safe Childbirth Checklist (SCC). The SCC encourages birth attendants to use EBPs known to prevent complications during the delivery process. By retrospectively applying the LAGO design to the BetterBirth study, we aimed to show how LAGO could have been used to adaptively satisfy an outcome goal of mean of performed EBPs greater than 0.8.

The BetterBirth study consisted of three stages. The first 2 stages were pilot trials, and stage 3 was a cluster randomized trial. The intervention package in stage 2 was modified based on feedback from stage 1 and adjusted again in stage 3. Stage 1 included 2 centers, stage 2 included 4 centers, and stage 3 included a control and an intervention arm with 15 centers each.

The outcome of interest we modeled was the proportion of EBPs performed out of all possible birth practices measured during each stage, and we modeled this as a continuous outcome. Births with 0 measured EBPs were excluded because whether EBPs were measured depended on the availability of researchers to document them. The number of EBPs measured at each stage were 14,19, and 18, and the average proportion of EBPs performed out of all possible birth practices measured was 0.33, 0.28, and 0.42 for the 3 stages, respectively.

We included the approximate monthly birth volume as a baseline center characteristic **z**. To avoid multicollinearity, we only considered 2 out of the 4 intervention components: the duration of the on-site intervention launch (in days) and the number of coaching visits after the initial intervention launch, truncated to 40 visits or less. Because the outcome of interest was the mean proportion of EBPs performed, we fit this LAGO study by a GLM with a logit link function to restrict the expected fractions to values between 0 and 1:

$$\begin{aligned} &logit\left\{E(Y_{ij} \mid \mathbf{A}_j = \mathbf{a}_j, \mathbf{X}_j = \mathbf{x}_j; \boldsymbol{\beta}, \mathbf{z}_j)\right\} \\ &= \beta_0 + \boldsymbol{\beta}_1^T \mathbf{a}_j + \boldsymbol{\beta}_2^T \mathbf{z}_j . \end{aligned}$$

Table 2 reports the estimated effects of the intervention package components after each stage, based on all data available at the end of that stage. The final analysis (last column) indicates that both the duration of the on-site intervention launch and the number of coaching visits had positive effects. The estimated effect of the number of coaching visits was also highly significant across the different stages, and both the 2 sample means test and the P degrees-of-freedom χ^2 test had P-values less than 0.001. Next, we estimated the optimal intervention package resulting in a mean of performed EBPs greater than $0.8(\theta = 0.8)$ while minimizing cost. Let x_1 be the launch duration (in days) and x_2 be the number of coaching visits. We used the linear cost function $C_1(\mathbf{x}) = 800x_1 + 170x_2$ (Nevo et al., 2021) and the cubic cost function $C_2(\mathbf{x}) = 1700x_1 - 950x_1^2 + 220x_1^3 + 380x_2 - 24x_2^2 + 0.6x_2^3$. The cubic cost function revised the linear cost function to include an economy of scale at lower values of the intervention components and by including prohibitive cost as the intervention components neared their upper limits. The constraints were set such that $x_1 \in [1,\ 5]$ and $x_2 \in [1,\ 40]$. For a center with an average birth volume ($z = 175$), the estimated optimal intervention package under the linear cost function was 5 days of launch duration and 31 coaching visits, with a total cost of $9270. With the cubic cost function, the estimated optimal intervention package was 3.97 days of launch duration and 35.50 coaching visits, with a total cost of $15629. The closest integer values for the estimated optimal intervention that leads to an estimated mean of performed EBPs greater than 0.8, while minimizing cost were 4 days of launch duration and 36 coaching visits.

9 DISCUSSION

The LAGO design allows for changes in the composition of the intervention package based on accumulating data from an ongoing trial at pre-specified stages. LAGO could help prevent failed trials by adapting and optimizing the intervention package composition while the trial is ongoing. LAGO is useful for implementation trials, pragmatic trials, and clinical trials of combination regimens. Using methods that are fundamentally different from those of Nevo et al. (2021), this article extends the LAGO design to trials with continuous outcomes. We anticipate that LAGO will be widely adopted in trials of combination implementation strategies.

While LAGO designs in their current form do not incorporate interim hypothesis testing, futility stops can be included in the LAGO design. If after stage k we fail to obtain any recommended intervention within the feasible bounds that is projected to yield adequate power based on a predefined minimal acceptable power level, the trial may be terminated early. Type I errors do not increase from futility stops, since there is no strong conclusion when a trial terminates early for this reason (Snapinn et al., 2006).

Current LAGO designs warrant discussion of potential causal inference challenges. First, although the algorithms for calculating recommended interventions are fixed, centers in large-scale multi-stage intervention studies often implement intervention packages in substantially different ways. This unplanned variation introduces a form of quasi-randomness that aids in estimating effects by mitigating multicollinearity. Addressing confounding by indication within the LAGO design remains an important area for future research. The model in equation (1) implicitly assumes that the $\mathbf{z}_j$ suffice to control for center effects. Additionally, in LAGO trials with control periods, time trends will either have to be absent (Nevo et al., 2021) or included in the outcome model.

While the current LAGO design focuses on center-level interventions, which aligns with the cluster-level structure commonly used in large-scale implementation trials, extending our methods to address individual-level interventions represents an important direction for future research, as it allows for more personalized optimal intervention strategies.

Since the estimating equations (3) considered in this article are similar to those in Balan and Schiopu-Kratina (2005), an alternative theoretical approach for showing asymptotic properties of $\hat{\boldsymbol{\beta}}$ might be possible using martingale theory (Balan and Schiopu-Kratina, 2005). Assuming that the residuals form a martingale difference sequence, it might be possible to apply a triangular Martingale Central Limit Theorem [similar to van der Laan (2008)] to develop theory for LAGO with both large and small sample sizes per stage. This approach might also be useful in developing theory for LAGO with time-to-event outcomes.

The LAGO design will be applied to the PULESA-Uganda trial (Semitala and Longenecker, 2021), which aims to improve HIV-hypertension care by identifying barriers, designing implementation strategies, and evaluating their effectiveness and economic sustainability in Uganda's Kampala and Wakiso districts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank Dr Katherine Semrau for her valuable contributions in explaining the BetterBirth study data.

FUNDING

This research was supported by NIH Grant HL167936.

DATA AVAILABILITY

The data that support the findings in this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Balan RM and Schiopu-Kratina I (2005) Asymptotic results with generalized estimating equations for longitudinal data. The Annals of Statistics, 33, 522–541.

Beidas RS, Dorsey S, Lewis CC, Lyon AR, Powell BJ, Purtle J et al. (2022) Promises and pitfalls in implementation science from the perspective of US-based researchers: learning from a pre-mortem. Implementation Science, 17, 55. [PubMed: 35964095]

Collins LM, Murphy SA and Strecher V (2007) The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. American Journal of Preventive Medicine, 32, S112–S118. [PubMed: 17466815]

Fahrmeir L and Tutz G (2013) Multivariate Statistical Modelling Based on Generalized Linear Models. New York, NY: Springer Science and Business Media.

FDA. (2016) Adaptive designs for medical device clinical studies: Guidance for industry and food and drug administration staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-designs-medical-device-clinical-studies. [Accessed 22 March 2023].

FDA. (2019). Adaptive design clinical trials for drugs and biologics guidance for industry.https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry. [Accessed 2 April 2023].

Fogel DB (2018) Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemporary Clinical Trials Communications, 11, 156–164. [PubMed: 30112460]

Hirschhorn LR, Semrau K, Kodkany B, Churchill R, Kapoor A, Spector J et al. (2015) Learning before leaping: integration of an adaptive study design process prior to initiation of BetterBirth, a large-scale randomized controlled trial in Uttar Pradesh, India. Implementation Science, 10, 1–9. [PubMed: 25567289]

Liang K-Y and Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika, 73, 13–22.

McCullagh P and Nelder JA (2019) Generalized Linear Models. New York, NY: Routledge.

Nevo D, Lok JJ and Spiegelman D (2021) Analysis of “Learn-As-youGO”(LAGO) studies. The Annals of Statistics, 49, 793–819. [PubMed: 35510045]

Scheffe H (1999) The Analysis of Variance, vol. 72. Hoboken, NJ: John Wiley and Sons.

Semitala FC and Longenecker CT (2021) PULESA-UGANDA-strengthening the blood pressure care and treatment cascade for Ugandans living with HIV-ImpLEmentation strategies to sAve lives. NHLBI. Grant Number: 1UG3HL154501–01. https://grantome.com/grant/NIH/UG3-HL154501-01. [Accessed 27 March 2023].

Semrau KE, Hirschhorn LR, Marx Delaney M, Singh VP, Saurastri R, Sharma N et al. (2017) Outcomes of a coaching-based WHO safe childbirth checklist program in India. New England Journal of Medicine, 377, 2313–2324. [PubMed: 29236628]

Shi C, Song R, Lu W and Fu B (2018) Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. Journal of the Royal Statistical Society Series B: Statistical Methodology, 80, 681–702. [PubMed: 30555269]

Snapinn S, Chen M-G, Jiang Q and Koutsoukos T (2006) Assessment of futility in clinical trials. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 5, 273–281.

Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N et al. (2014) Adult cancer clinical trials that fail to complete: An epidemic? Journal of the National Cancer Institute, 106, dju229. [PubMed: 25190726]

van der Laan MJ (2008) The construction and analysis of adaptive group sequential designs. Technical Report 232. Division of Biostatistics, UC Berkeley.

van der Vaart AW (2000) Asymptotic Statistics, vol. 3. Cambridge, UK: Cambridge University Press.

TABLE 1

Selected simulation study results for simulation 1 scenarios 1a and 1b with a linear cost function.

(a) Simulation study results for individual package component effects with a linear cost function

$\beta^* = (\beta_{11}^*, \beta_{12}^*)$	$n_j^{(1)}$	$n_j^{(2)}$	J	%RelBias	$\hat{\beta}_{11}\frac{SE}{EMP.SD}(\times 100)$	CP95	%RelBias	$\hat{\beta}_{12}\frac{SE}{EMP.SD}(\times 100)$	CP95
Scenario 1a $J_1 = J_2 = J$									
(0.1863, 0.15)	50	100	6	1.64	84.8	96.1	−0.46	76.1	95.0
			10	1.78	90.6	94.9	−0.55	79.3	94.9
			20	1.54	97.6	95.1	−0.37	91.2	95.2
		200	6	1.44	77.9	95.2	−0.42	70.9	95.5
			10	1.91	86.1	95.4	−0.49	73.6	95.5
			20	1.14	97.8	95.4	−0.26	88.2	94.8
Scenario 1b $(J_1 = 6, J_2 = 12)$									
(0.1863, 0.15)	50	200		0.04	85.9	95.0	−0.09	75.0	94.8
(0.1, 0.2133)	50	200		0.56	93.7	95.8	0.14	84.3	95.7

(b) Simulation study results for estimated optimal intervention with a linear cost function

$\beta^* = (\beta_{11}^*, \beta_{12}^*)$	x^{opt}	$n_j^{(1)}$	$n_j^{(2)}$	Stage 1			Stage 2/LAGO optimized		
				Bias of $\hat{x}_1^{opt}(\times 100)$	Bias of $\hat{x}_2^{opt}(\times 100)$	rMSE (×100)	Bias of $\hat{x}_1^{opt}(\times 100)$	Bias of $\hat{x}_2^{opt}(\times 100)$	rMSE (×100)
Scenario 1a $(J_1 = J_2 = 20)$									
(0.1863, 0.15)	(1,8)	50	100	15.2	96.9	115.0	−0.27	0.01	35.3
			500				−0.70	0.00	28.2
		100	100	14.0	51.6	94.1	0.11	0.01	32.7
			500				−0.25	0.00	24.5
Scenario 1b $(J_1 = 6, J_2 = 12)$									
(0.1863, 0.15)	(1,8)	50	200	20.7	341.5	191.1	−1.5	1.5	52.3
(0.1, 0.2133)	(0, 6.5)	50	200	−46.2	239.2	172.9	−0.1	0.1	39.2

(c) Simulation study results for estimated optimal intervention, confidence set, and confidence band with a linear cost function

$\beta^* = (\beta_{11}^*, \beta_{12}^*)$	x^{opt}	$n_j^{(1)}$	$n_j^{(2)}$	**MeanOpt1 (Q2.5,Q97.5)**	**MeanOpt2 (Q2.5,Q97.5)**	**SetCP95 %**	**SetPerc %**	**Bands CP95%**
Scenario 1a ($J_1 = J_2 = 20$)								
(0.1863, 0.15)	(1,8)	50	100	(0.655, 0.818)	(0.789, 0.811)	95.3	5.5	95.2
		500		(0.794, 0.808)	95.1	3.7	95.7	
	100	100	(0.708, 0.816)	(0.791, 0.808)	94.4	4.4	95.1	
		500		(0.795, 0.805)	95.2	2.8	95.8	
Scenario 1b ($J_1 = 6, J_2 = 12$)								
(0.1863, 0.15)	(1,8)	50	200	(0.529, 0.826)	(0.773, 0.826)	94.5	9.9	95.4
(0.1, 0.2133)	(0, 6.5)	50	200	(0.515, 0.857)	(0.783, 0.818)	95.7	12.6	95.7

$n_j^{(1)}$, number of patients in center j at stage 1, $n_j^{(2)}$, number of patients in center j at stage 2 . J, number of centers for each stage.

%RelBias, percent relative bias $100(\hat{\beta} - \beta^\star)/\beta^\star$.

SE, mean estimated standard error, EMP.SD, empirical standard deviation.

CP95, empirical coverage rate of 95% confidence intervals.

Bias of $\hat{x}_1^{opt}$, bias of the first component of the estimated optimal intervention, Bias of $\hat{x}_2^{opt}$, bias of the second component of the estimated optimal intervention.

rMSE, root of mean squared errors, $\left\{\text{mean}\left(\|\hat{\mathbf{x}}^{\text{opt}} - \mathbf{x}^{\text{opt}}\|^2\right)\right\}^{1/2}$, mean is taken over simulation iterations.

MeanOpt1, mean outcome under the stage 2 recommended intervention, calculated using true coefficient values; MeanOpt2, mean outcome under the final estimated optimal intervention based on all data, calculated using true coefficient values.

Q2.5 and Q97.5: 2.5% and 97.5% quantiles.

SetCP95%, empirical coverage percentage of confidence set for the optimal intervention. SetPerc%, mean percentage of the size of the confidence set as a percent of the total sample space. BandsCP95%, empirical coverage of 95% confidence band.

TABLE 2

The BetterBirth study: package component effect estimates, 95% confidence intervals after each stage, and estimated optimal intervention package after each stage.

	Stage 1	**Stage 1–2**	**Stage 1–3**
	$n^{(1)} = 113$	$n^{(1)} + n^{(2)} = 2256$	$n^{(1)} + n^{(2)} + n^{(3)} = 7342$
	$\hat{\beta}$ **(95% CI)**	$\hat{\beta}$ **(95% CI)**	$\hat{\beta}$ **(95% CI)**
Intercept	2.72 (1.20, 4.25)	−0.61 (−0.69, −0.53)	−0.138 (−0.156, −0.120)
Launch duration (days)	−0.09 (−0.33, 0.14)	−0.003 (−0.11, 0.10)	0.17 (0.11, 0.22)
Coaching visits (per 5 visits)	0.90 (0.79, 1.01)	0.32 (0.31, 0.34)	0.172 (0.167, 0.176)
Birth volume (monthly, per 100)	−3.39 (−4.59, −2.18)	−0.166 (−0.180, −0.153)	−0.202 (−0.210, −0.195)
$\hat{\mathbf{x}}^{opt}$ (using the linear cost function)	(1, 16)	(1, 36)	(5, 31)
$\hat{\mathbf{x}}^{opt}$ (using the cubic cost function)	(1, 15.66)	(1, 35.38)	(3.97, 35.50)

CI, based on sandwich estimator for $\mathrm{VAR}\left(\hat{\boldsymbol{\beta}}\right)$ (see Theorem 2).

For (optimal) interventions, the first component is launch duration and the second component is number of coaching visits.

The optimal intervention reported is for a center with an average birth volume ($z = 175$).