Supplementary materials for: Analysis of "Learn-As-You-Go" Studies

Daniel Nevo, Judith J. Lok and Donna Spiegelman

1 Description of what is included in the supplementary materials

This document complements the paper by Nevo, Lok and Spiegelman. In Section 2 we prove that the estimator solving Equation (2.3) in the main paper is a partial maximum likelihood estimator. In Section 3, we outline how to generalize the proofs of Theorems 2.1 and 2.2 to a general number of stages K. In Section 4, we show that the proposed confidence bands for the success probabilities have a simultaneous asymptotic coverage rate of at least 95%. In Section 5.1, we describe the function g that was used in the simulation study to choose the stage-2 recommended interventions $\boldsymbol{X}_j^{(2,n_1)}$ based on the stage 1 data and the cost function. In Section 5.2.1, we present additional simulation results. In Section 6, we present confidence bands for the probability of oxytocin administration based on the BetterBirth data analysis.

2 The proposed estimator is a maximum partial likelihood estimator

The probability function of the entire data is $Pr(\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{X}}^{(2,n_1)}, \overline{\boldsymbol{A}}^{(2,n_1)}, \overline{\boldsymbol{Y}}^{(2,n_1)}, |\overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)})$, and can be decomposed in the following way:

$$
\begin{aligned}
&Pr(\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{X}}^{(2,n_1)}, \overline{\boldsymbol{A}}^{(2,n_1)}, \overline{\boldsymbol{Y}}^{(2,n_1)}, |\overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \\
&= Pr(\overline{\boldsymbol{Y}}^{(1)}|\overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \times Pr(\overline{\boldsymbol{X}}^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \\
&\times Pr(\overline{\boldsymbol{A}}^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{X}}^{(2,n_1)}, \overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \times Pr(\overline{\boldsymbol{Y}}^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{A}}^{(2,n_1)}, \overline{\boldsymbol{X}}^{(2,n_1)}, \overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \quad \text{(A1)} \\
&= Pr(\overline{\boldsymbol{Y}}^{(1)}|\overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{a}}^{(1)},) Pr(\overline{\boldsymbol{X}}^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(2)}) \times Pr(\overline{\boldsymbol{A}}^{(2,n_1)}|\overline{\boldsymbol{X}}^{(2,n_1)}) \\
&\times Pr(\overline{\boldsymbol{Y}}^{(2,n_1)}|\overline{\boldsymbol{A}}^{(2,n_1)}, \overline{\boldsymbol{z}}^{(2)}) \quad \text{(A2)} \\
&= \left[\prod_{j=1}^{J_1} Pr(\boldsymbol{Y}_j^{(1)}|\boldsymbol{z}_j^{(1)}, \boldsymbol{a}_j^{(1)})\right] \times \left[\prod_{j=1}^{J_2} Pr(\boldsymbol{X}_j^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{z}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \boldsymbol{z}_j^{(2)})\right] \\
&\times \left[\prod_{j=1}^{J_2} Pr(\boldsymbol{A}_j^{(2,n_1)}|\boldsymbol{X}_j^{(2,n_1)})\right] \times \left[\prod_{j=1}^{J_2} Pr(\boldsymbol{Y}_j^{(2,n_1)}|\boldsymbol{A}_j^{(2,n_1)}, \boldsymbol{z}_j^{(2)})\right] \quad \text{(A3)} \\
&= \left[\prod_{j=1}^{J_1} Pr(\boldsymbol{Y}_j^{(1)}|\boldsymbol{z}_j^{(1)}, \boldsymbol{a}_j^{(1)})\right] \times \left[\prod_{j=1}^{J_2} Pr(\boldsymbol{Y}_j^{(2,n_1)}|\boldsymbol{A}_j^{(2,n_1)}, \boldsymbol{z}_j^{(2)})\right] \quad \text{(A4)} \\
&= \left[\prod_{j=1}^{J_1}\prod_{i=1}^{n_{j1}} (p_{\boldsymbol{a}_j^{(1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(1)}))^{Y_{ij}^{(1)}} (1 - p_{\boldsymbol{a}_j^{(1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(1)}))^{1-Y_{ij}^{(1)}}\right] \\
&\times \left[\prod_{j=1}^{J_2}\prod_{i=1}^{n_{j2}} (p_{\boldsymbol{A}_j^{(2,n_1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(2)}))^{Y_{ij}^{(2,n_1)}} (1 - p_{\boldsymbol{A}_j^{(2,n_1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(2)}))^{1-Y_{ij}^{(2,n_1)}}\right].
\end{aligned}
$$

In (A1), the joint distribution is decomposed. In (A2), we use Assumption 2.1 and the assumption that given $\boldsymbol{A}$, $\boldsymbol{Y}$ and $\boldsymbol{X}$ are independent. (A3) is justified by the assumption of independence between centers. (A4) is justified by the deterministic dependence of $\boldsymbol{X}_j^{(2,n_1)}$ on the stage 1 data and $\boldsymbol{z}_j^{(2)}$, so $Pr(\boldsymbol{X}_j^{(2,n_1)}|\overline{\boldsymbol{Y}}^{(1)}, \overline{\boldsymbol{a}}^{(1)}, \overline{\boldsymbol{z}}^{(1)}, \boldsymbol{z}_j^{(2)}) = 1$, and by the assumption that $Pr(\boldsymbol{A}_j^{(2,n_1)}|\boldsymbol{X}_j^{(2,n_1)}) = 1$ since $\boldsymbol{A}_j^{(2,n_1)} = h_j^{(1)}(\boldsymbol{X}_j^{(2,n_1)})$, h is a deterministic function.

Therefore, a partial likelihood for $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \left[\prod_{j=1}^{J_1} \prod_{i=1}^{n_{j1}} (p_{\boldsymbol{a}_j^{(1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(1)}))^{Y_{ij}^{(1)}} (1 - p_{\boldsymbol{a}_j^{(1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(1)}))^{1-Y_{ij}^{(1)}} \right] \times \left[\prod_{j=1}^{J_2} \prod_{i=1}^{n_{j2}} (p_{\boldsymbol{A}_j^{(2,n_1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(2)}))^{Y_{ij}^{(2,n_1)}} (1 - p_{\boldsymbol{A}_j^{(2,n_1)}}(\boldsymbol{\beta}; \boldsymbol{z}_j^{(2)}))^{1-Y_{ij}^{(2,n_1)}} \right].$$

By taking derivatives with respect to $\boldsymbol{\beta}$, it can be seen that maximizing $L(\boldsymbol{\beta})$ is equivalent to solving $U(\boldsymbol{\beta})$=0, with $U(\boldsymbol{\beta})$ defined by Equation (2.3) in the main text.

3 Extension to a general number of stages

In this section, we consider the case of K stages, with $K \geq 3$. We start with adjusting the notation from a two-stage LAGO design to a K-stage LAGO design. The notation for stage 1 is the same as in the main text. Let J_k number of centers at stage k, and let $\boldsymbol{z}_j^{(k)}$ be the center-specific characteristics in center j of stage k. Similar to the definitions in the main text, let $\boldsymbol{X}_j^{(k,\overline{n}_{k-1})}, j' = 1, ..., J_k$, be the recommended intervention in center j of stage k. $\boldsymbol{X}_j^{(k,\overline{n}_{k-1})}$ is chosen according to functions g_k of the entire data up to stage $k-1$, including the stage $k-1$ data. The actual intervention in center j of stage k is $\boldsymbol{A}_j^{(k,\overline{n}_{k-1})} = h_j^{(k)}(\boldsymbol{X}_j^{(k,\overline{n}_{k-1})})$, where $h_j^{(k)}$ is a deterministic center-specific continuous function from $\mathcal{X}$ to $\mathcal{X}$ that describes how center j of stage k decides on the intervention based on the recommendation $\boldsymbol{X}_j^{(k,\overline{n}_{k-1})}$.

Once the recommended intervention $\boldsymbol{X}_j^{(k,\overline{n}_{k-1})}$ is determined, stage k outcomes are collected under the intervention value $\boldsymbol{A}_j^{(k,\overline{n}_{k-1})}$. Let $\boldsymbol{Y}_j^{(k,\overline{n}_{k-1})} = (Y_{1j}^{(k,\overline{n}_{k-1})}, ..., Y_{n_{kj}j}^{(k,\overline{n}_{k-1})})$ be the vector of outcomes in center j of stage k. Let $\overline{\boldsymbol{X}}^{(k,\overline{n}_{k-1})} = (\boldsymbol{x}^{(1)}, \boldsymbol{X}^{(2,n_1)}, ..., \boldsymbol{X}^{(k,\overline{n}_{k-1})})$, $\overline{\boldsymbol{A}}^{(k,\overline{n}_{k-1})} = (\boldsymbol{a}^{(1)}, \boldsymbol{A}^{(2,n_1)}, ..., \boldsymbol{A}^{(k,\overline{n}_{k-1})})$ and $\overline{\boldsymbol{Y}}^{(k,\overline{n}_{k-1})} = (\boldsymbol{Y}^{(1)}, ..., \boldsymbol{Y}^{(k,\overline{n}_{k-1})})$ be the recommended interventions, actual interventions and outcomes until stage k, respectively. We assume that conditionally on the stage k actual interventions, $(\overline{\boldsymbol{A}}^{(k,\overline{n}_{k-1})}, \overline{\boldsymbol{Y}}^{(k,\overline{n}_{k-1})})$ are independent of $\overline{\boldsymbol{X}}^{(k,\overline{n}_{k-1})}$ and the data from previous stages. Similar to the case of $K = 2$, we assume the distribution of $\overline{\boldsymbol{Y}}^{(k,\overline{n}_{k-1})}$ follows the logistic regression model (1) given in the main paper.

The two main assumptions in the main text are adjusted for general number of stages $K > 2$, in the following way:

Assumption A1 *Conditionally on $\bar{\boldsymbol{X}}^{(2,n_1)}$, $(\bar{\boldsymbol{A}}^{(2,n_1)}, \bar{\boldsymbol{Y}}^{(2,n_1)})$ are independent of the stage 1 data $(\bar{\boldsymbol{a}}^{(1)}, \bar{\boldsymbol{Y}}^{(1)})$.*

Assumption A2 *For all $k = 3, ..., K$, conditionally on $\overline{\boldsymbol{X}}^{(k,\overline{n}_{k-1})}$, $(\overline{\boldsymbol{A}}^{(k,\overline{n}_{k-1})}, \overline{\boldsymbol{Y}}^{(k,\overline{n}_{k-1})})$ are independent of the data from the first $k-1$ stages, $(\overline{\boldsymbol{A}}^{(k-1,\overline{n}_{k-2})}, \overline{\boldsymbol{Y}}^{(k-1,\overline{n}_{k-2})})$.*

Assumption A3 *For all $j = 1, ..., J_k$ and $k = 1, ..., K$, the recommended intervention in center j of stage k, $\boldsymbol{X}_j^{(k,\overline{n}_{k-1})}$, converges in probability to a center-specific limit $\boldsymbol{x}_j^{(k)}$.*

Assumption A1 is identical to Assumption 2.1 in the paper. Assumption A2 extends Assumption 2.1 in the paper for later stages. This is done because the stage 1 intervention $\boldsymbol{a}^{(1)}$ is nonrandom. Assumption A3 extends Assumption 2.2 in the main text to the scenario with $K > 2$ stages.

Let $\boldsymbol{a}_j^{(k)} = h_j^{(k)}(\boldsymbol{x}_j^{(k)})$ be the probability limit of $\boldsymbol{A}_j^{(k,\overline{n}_{k-1})}$. We now outline the adjustments needed in the proofs presented in Appendices A.1 and A.2 of the main paper to extend the results to a general number of stages K. Consistency can be proved as before in the following way. The score function for K stages has expectation zero at the true $\boldsymbol{\beta}^\star$ by considering terms for the different stages separately. Covariances between terms for stage k and k' are again 0, as can be seen by conditioning on the intervention at stage $\max(k, k')$. Consistency then follows as for the case with 2 stages.

For asymptotic normality, by arguments similar to those in the proof of Theorem 2.2, we obtain that the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$ is the same as the asymptotic distribution of the K-stages analogue of Equation (2.5) in the main text, i.e.,

$$\begin{aligned}[I(\boldsymbol{\beta}^\star)]^{-1}\frac{1}{\sqrt{n}}\Bigg[&\sum_{j=1}^{J_1}\sum_{i=1}^{n_{j1}}\begin{pmatrix}1\\ \boldsymbol{a}_j^{(1)}\\ \boldsymbol{z}_j^{(1)}\end{pmatrix}\left(Y_{ij}^{(1)} - p_{\boldsymbol{a}_j^{(1)}}(\boldsymbol{\beta}^\star;\boldsymbol{z}_j^{(1)})\right)\\ &+\sum_{k=2}^{K}\sum_{j=1}^{J_2}\sum_{i=1}^{n_{jk}}\begin{pmatrix}1\\ \boldsymbol{A}_j^{(k,\overline{n}_{k-1})}\\ \boldsymbol{z}_j^{(k)}\end{pmatrix}\left(Y_{ij}^{(k,\overline{n}_{k-1})} - p_{\boldsymbol{A}_j^{(k,\overline{n}_{k-1})}}(\boldsymbol{\beta}^\star;\boldsymbol{z}_j^{(k)})\right)\Bigg],\end{aligned} \tag{A5}$$

where $I(\boldsymbol{\beta}^\star)$ is redefined for a general number of stages.

The appropriate adjustment of the coupling argument uses backward induction, as follows. First, we couple the stage K outcomes with outcomes $\boldsymbol{Y}_j^{(K)}$ had stage K center j had an actual intervention $\boldsymbol{a}_j^{(K)}$, resulting in $\tilde{\boldsymbol{Y}}_j^{(K,\overline{n}_{K-1})}$ as in Equations (2.6) and (A.14) in the main text, with 2 replaced by K. As in the coupling for stage 2 outcomes in the $K = 2$ scenario, replacing $\boldsymbol{Y}_j^{(K,\overline{n}_{K-1})}$ by $\tilde{\boldsymbol{Y}}_j^{(K,\overline{n}_{K-1})}$ does not change the distribution of the resulting expression. The difference between the resulting expression with $\tilde{\boldsymbol{Y}}_j^{(K,\overline{n}_{K-1})}$ and $\boldsymbol{A}_j^{(K,\overline{n}_{K-1})}$ and the same expression with $\boldsymbol{Y}_j^{(K)}$ and $\boldsymbol{a}_j^{(K)}$ converges in probability to zero, as for stage 2 before, since $\boldsymbol{A}_j^{(K,\overline{n}_{K-1})} \xrightarrow{P} \boldsymbol{a}_j^{(K)}$. Thus, we can replace the $\tilde{\boldsymbol{Y}}^{(K,\overline{n}_{K-1})}$ by $\boldsymbol{Y}_j^{(K)}$ and $\boldsymbol{A}_j^{(K,\overline{n}_{K-1})}$ by $\boldsymbol{a}_j^{(K)}$ without changing the asymptotic distribution. The resulting stage K term is independent of all other terms. Then, we can continue with the term for stage $K-1$ and use backward induction to show that a similar replacement can be done for all stages. Upon replacing in (A5) all $\boldsymbol{Y}^{(k,\overline{n}_{k-1})}$ by $\boldsymbol{Y}^{(k)}$ and all $\boldsymbol{A}_j^{(k,\overline{n}_{k-1})}$ by $\boldsymbol{a}_j^{(k)}$ ($j_1, ..., J_k$ and $k = 2, ..., K$), the resulting expression is a sum of K independent terms, and classical logistic regression theory leads to the result of Theorem 2.2.

4 Confidence bands for the success probabilities

Our proposed procedure is similar to the one described by Scheffé (1959), and uses our asymptotic normality results about $\boldsymbol{\beta}$. We construct confidence bands for $p_{\boldsymbol{x}}(\boldsymbol{\beta};\tilde{\boldsymbol{z}})$, for all $\boldsymbol{x} \in \mathcal{X}$ and for a fixed $\boldsymbol{z} = \tilde{\boldsymbol{z}}$. Consider confidence bands of the form

$$Pr\left[CB_{p_{\boldsymbol{x}}} \ni p_{\boldsymbol{x}}(\boldsymbol{\beta}^\star;\tilde{\boldsymbol{z}},\ \text{for all } \boldsymbol{x} \in \mathcal{X}\right] = 1 - \alpha.$$

Since

$$Pr[CB_{p_{\boldsymbol{x}}} \ni p_{\boldsymbol{x}}(\boldsymbol{\beta}^\star;\tilde{\boldsymbol{z}})] = Pr[\text{logit}\, CB_{p_{\boldsymbol{x}}} \ni \text{logit}\, p_{\boldsymbol{x}}(\boldsymbol{\beta}^\star;\tilde{\boldsymbol{z}})] = Pr[\text{logit}\, CB_{p_{\boldsymbol{x}}} \ni (\beta_0^\star + \boldsymbol{x}^T\boldsymbol{\beta}_1^\star + \tilde{\boldsymbol{z}}^T\boldsymbol{\beta}_2^\star)],$$

we can construct confidence bands for $\beta_0^\star + {\boldsymbol{\beta}_1^\star}^T\boldsymbol{x} + {\boldsymbol{\beta}_2^\star}^T\tilde{\boldsymbol{z}}$ (for all $\boldsymbol{x} \in \mathcal{X}$), denoted by $CB_{\boldsymbol{x}}$, and obtain confidence bands for $p_{\boldsymbol{x}}$ by setting $CB_{p_{\boldsymbol{x}}} = \text{expit}(CB_{\boldsymbol{x}})$. To this end, we will find $C_\alpha(\boldsymbol{x})$ such that

$$Pr\left[[(\hat{\beta}_0 - \beta_0^\star) + (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^\star)^T\boldsymbol{x} + (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^\star)^T\tilde{\boldsymbol{z}}]^2 \le C_\alpha(\boldsymbol{x}),\ \text{for all } \boldsymbol{x} \in \mathcal{X}\right] = 1 - \alpha.$$

A special case that is convenient to work with is

$$\begin{aligned}1 - \alpha =& Pr\left[\frac{[(\hat{\beta}_0 - \beta_0^\star) + (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^\star)^T\boldsymbol{x} + (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^\star)^T\tilde{\boldsymbol{z}}]^2}{(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)I^{-1}(\boldsymbol{\beta}^\star)(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T} \le \tilde{C}_\alpha,\ \text{for all } \boldsymbol{x} \in \mathcal{X}\right]\\ =& Pr\left[\max_{\boldsymbol{x}\in\mathcal{X}}\left\{\frac{[(\hat{\beta}_0 - \beta_0^\star) + (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^\star)^T\boldsymbol{x} + (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^\star)^T\tilde{\boldsymbol{z}}]^2}{(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)I^{-1}(\boldsymbol{\beta}^\star)(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T}\right\} \le \tilde{C}_\alpha\right].\end{aligned}$$

Now, by setting $\boldsymbol{u} = I^{-1/2}(\boldsymbol{\beta}^\star)(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T$,

$$\max_{\boldsymbol{x}\in\mathcal{X}} \left\{ \frac{[(\hat\beta_0 - \beta_0^\star) + (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^\star)^T\boldsymbol{x} + (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^\star)^T\tilde{\boldsymbol{z}}]^2}{(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)I^{-1}(\boldsymbol{\beta}^\star)(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T} \right\} = \max_{\boldsymbol{u}\in\widetilde{\mathcal{X}}} \left\{ \frac{[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)^T I^{1/2}(\boldsymbol{\beta}^\star)\boldsymbol{u}]^2}{\boldsymbol{u}^T\boldsymbol{u}} \right\}, \tag{A6}$$

where $\widetilde{\mathcal{X}}$ is obtained by mapping each point $\boldsymbol{x} \in \mathcal{X}$ to $\boldsymbol{u} = I^{-1/2}(\boldsymbol{\beta}^\star)(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T$. By the Cauchy—Schwarz inequality, the maximum in the right hand side of (A6) is lower or equal of the term when taking $\boldsymbol{u} = I^{-1/2}(\boldsymbol{\beta}^\star)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)$. That is,

$$\max_{\boldsymbol{u}\in\widetilde{\mathcal{X}}} \left\{ \frac{[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)^T I^{1/2}(\boldsymbol{\beta}^\star)\boldsymbol{u}]^2}{\boldsymbol{u}^T\boldsymbol{u}} \right\} \le ||(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)^T I^{1/2}(\boldsymbol{\beta}^\star)||^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)^T I(\boldsymbol{\beta}^\star)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)$$

which, when multiplied by n, has an asymptotic Chi-square distribution with $p + q + 1$ degrees of freedom by the asymptotic normality we proved in Theorem 2.2. Thus, $\tilde{C}_\alpha = n^{-1}\chi^2_{(1-\alpha,p+q+1)}$, where $\chi^2_{(1-\alpha,p+q+1)}$ is the $1-\alpha$ quantile of Chi-square distribution with $p+q+1$ degrees of freedom. We conclude that conservative 95% confidence bands for logit $p_{\boldsymbol{x}}(\boldsymbol{\beta}^\star; \tilde{\boldsymbol{z}})$ are given by

$$CB_{\boldsymbol{x}} = \left[(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)\hat{\boldsymbol{\beta}} - \sqrt{\chi^2_{0.95,p+q+1}}\sigma(\hat{\boldsymbol{\beta}}; \boldsymbol{x}, \tilde{\boldsymbol{z}}), \qquad (1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)\hat{\boldsymbol{\beta}} + \sqrt{\chi^2_{0.95,p+q+1}}\sigma(\hat{\boldsymbol{\beta}}; \boldsymbol{x}, \tilde{\boldsymbol{z}})\right],$$

with $\sigma^2(\hat{\boldsymbol{\beta}}; \boldsymbol{x}, \tilde{\boldsymbol{z}}) = (1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)n^{-1}\hat{I}^{-1}(\hat{\boldsymbol{\beta}})(1\ \boldsymbol{x}^T\ \tilde{\boldsymbol{z}}^T)^T$ and $\chi^2_{0.95,p+q+1}$ being the 95% quantile of a χ^2_{p+q+1} distribution.

5 Additional information on the simulation study

In this section, we present further details about and results from the simulation study. In Section 5.1, we present the function g we used in the simulations to decide on the stage 2 recommended interventions $\boldsymbol{X}_j^{(2,n_1)}$ based on the stage 1 data. In Section 5.2, we present additional simulation results.

5.1 Determination of the recommended intervention in the simulation study

We present the function we used in the simulation study to determine the stage 2 recommended intervention. The simulation study did not have an intercept β_0. Consider a center with a center-specific covariate $z_j^{(2)} = \tilde{z}$ (in the simulations z is a single center characteristic). Let $\hat{\boldsymbol{\beta}}^{(1)} = (\hat\beta_{11}^{(1)}, \hat\beta_{12}^{(1)}, \hat\beta_2^{(1)})$ be the stage 1 estimator of $\boldsymbol{\beta}$, where $\hat\beta_{11}^{(1)}$ and $\hat\beta_{12}^{(1)}$ are the estimated individual component effects of the intervention and $\hat\beta_2^{(1)}$ is the estimated effect of z.

First, we check whether the target value $\tilde{p}$ can be reached for center j with $z_j^{(2)} = \tilde{z}$, using the estimated $\hat{\boldsymbol{\beta}}^{(1)}$. For example, if $\hat\beta_{11}^{(1)}$ and $\hat\beta_{12}^{(1)}$ are positive, we check if

$$\hat\beta_{11}^{(1)}\mathcal{U}_1 + \hat\beta_{12}^{(1)}\mathcal{U}_2 + \hat\beta_2^{(1)}\tilde{z} \ge \text{logit}(\tilde{p}), \tag{A7}$$

where $\mathcal{U}_1$ and $\mathcal{U}_2$ are the upper values of the intervention $\boldsymbol{X}$ as described in Section 2.1 of the main text. If there exists a solution to the optimization problem given by Equation (2.2) in the main text, with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}^{(1)}$, we set $\boldsymbol{X}_j^{(2,n_1)} = \hat{\boldsymbol{x}}_j^{opt,(2,n_1)}$ where $\hat{\boldsymbol{x}}_j^{opt,(2,n_1)}$ is the estimated optimal intervention based on the stage 1 data and obtained as described in Sections 2.1 and 2.5 of the main text.

However, if (7) is not satisfied, this means that $\tilde{p}$ cannot be reached under the currently estimated $\hat{\boldsymbol{\beta}}^{(1)}$ vector. Thus a more careful choice of $\boldsymbol{X}_j^{(2,n_1)}$ is needed. Let $\widetilde{X}_{11}$ be the solution of (now $\hat\beta_{11}^{(1)}$ is fixed)

$$\hat\beta_{11}^{(1)}\widetilde{X}_{11} + I\{\hat\beta_{12}^{(1)} > 0\}\hat\beta_{12}^{(1)}\mathcal{U}_2 + I\{\hat\beta_{12}^{(1)} < 0\}\hat\beta_{12}^{(1)}\mathcal{L}_2 + \hat\beta_2^{(1)}\tilde{z} = \text{logit}(\tilde{p}).$$

That is, $\widetilde{X}_{11}$ is the value of the first component of the package such that $\tilde{p}$ is achieved, regardless of the constraint that $X_1 \in [\mathcal{L}_1, \mathcal{U}_1]$. The dashed blue line in Figure 1 shows $\widetilde{X}_{j1}$. However, since values larger than $\mathcal{U}_1$ are impossible or unfeasible to implement (we are not allowed to set $\boldsymbol{X}_{j1}^{(2,n_1)} > \mathcal{U}_1$), $\widetilde{X}_{11}$ is of a little interest for small values of $\hat{\beta}_{11}^{(1)}$. Instead, we use $g(\hat{\boldsymbol{\beta}})$, which we now describe; $g(\hat{\boldsymbol{\beta}})$ is depicted by the red straight line in Figure 1.

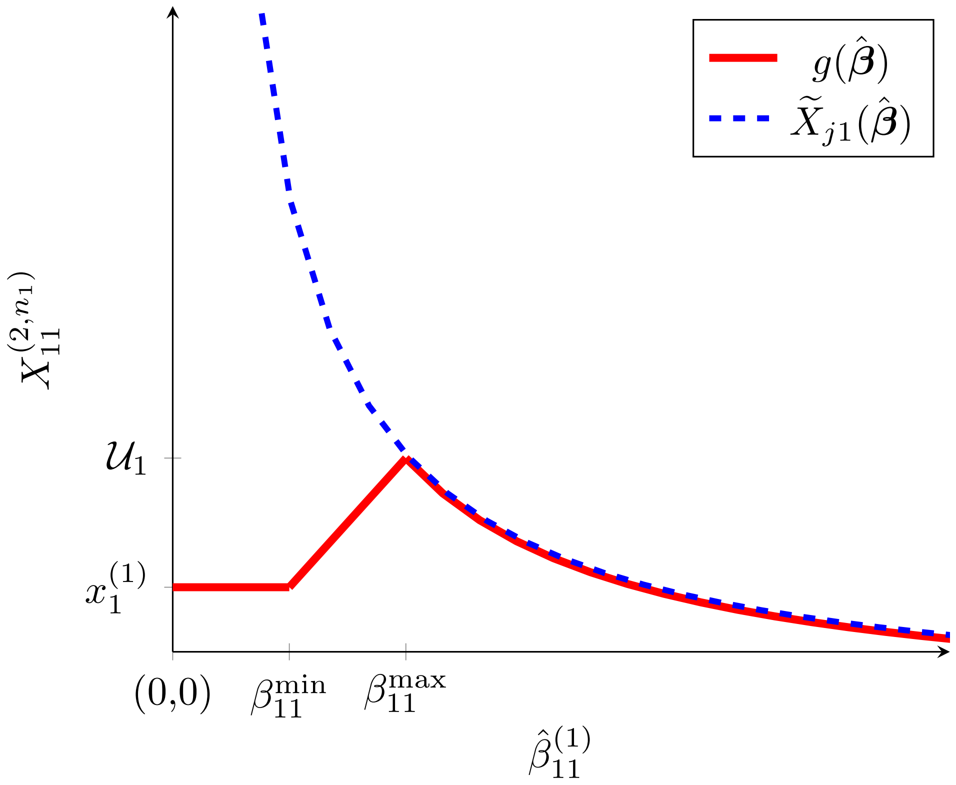


Figure 1: Illustration of the function g used in the simulation studies to choose $\boldsymbol{X}^{(2,n_1)}$. The x-axis depicts possible values of $\hat{\beta}_{11}$ and the y-axis depict the values of the first component of the recommended intervention in stage 2. The dashed blue line is the value of the optimal intervention, based on the stage 1 estimator, if the first package component was not bounded by $\mathcal{U}_1$. The solid red line is the function we used in the simulations to choose the stage 2 recommended intervention. For simplicity of presentation, the figure presents settings with $\hat{\beta}_{11}^{(1)}/c_1 > \hat{\beta}_{12}^{(1)}/c_2$. The figure assumes that the lower bound of X_1 equals $\mathcal{L}_1 = 0$.

The function $g(\hat{\boldsymbol{\beta}})$ we have used in the simulation study to decide on the stage 2 recommended intervention when there is no solution to the optimization problem works on each component of the intervention package separately, while the other component is held fixed at its maximal value if its effect estimate is positive and at its minimal value if its estimated effect is negative. The procedure is illustrated in Figure 1. Let $\beta_{11}^{\max}$ be the solution of

$$\beta_{11}^{\max}\mathcal{U}_1 + I\{\hat{\beta}_{12}^{(1)} > 0\}\hat{\beta}_{12}^{(1)}\mathcal{U}_2 + I\{\hat{\beta}_{12}^{(1)} < 0\}\hat{\beta}_{12}^{(1)}\mathcal{L}_2 + \hat{\beta}_2^{(1)}\tilde{z} = \text{logit}(\tilde{p}).$$

Because there is no solution to the optimization problem, $\beta_{11}^{\max} < \hat{\beta}_{11}^{(1)}$. Let $\beta_{11}^{\min}$ be some value lower than $\beta_{11}^{\max}$; we took $\beta_{11}^{\min} = \beta_{11}^{\max}/2$. Recall that $\boldsymbol{x}^{(1)}$ is the pre-specified stage 1 recommended intervention. Set the first component of $\boldsymbol{X}_j^{(2,n_1)}$, $X_{j1}^{(2,n_1)}$, to

$$X_{j1}^{(2,n_1)} = \begin{cases} x_1^{(1)} & \text{if } \hat{\beta}_{11}^{(1)} \le \beta_{11}^{\min} \\ x_1^{(1)} + \frac{\mathcal{U}_1 - x_1^{(1)}}{\beta_{11}^{\max} - \beta_{11}^{\min}}(\hat{\beta}_{11}^{(1)} - \beta_{11}^{\min}) & \text{if } \beta_{11}^{\min} < \hat{\beta}_{11}^{(1)} \le \beta_{11}^{\max}. \end{cases}$$

In words, keep $X_1^{(2,n_1)} = x_1^{(1)}$ if $\hat{\beta}_{11}^{(1)}$ is very small, assuming that either this component is ineffective or that more data are needed. As $\hat{\beta}_{11}^{(1)}$ increases, we increase $X_{j1}^{(2,n_1)}$. This also ensures that this stage 2 recommended intervention $X_{j1}^{(2,n_1)}$ is continuous in $\hat{\beta}_{11}^{(1)}$, as needed for Assumption 2.2 to hold.

The procedure described above is repeated for the second component of the intervention (now fixing the first intervention component at its minimal or maximal value). The procedure of choosing the stage 2 recommended intervention $\boldsymbol{X}_j^{(2,n_1)}$ is for each value of the stage 2 centers, based on their center characteristics $\tilde{z}_j^{(2)}$.

5.2 Additional simulation results

In this section, we present the following results:

- A summary of the results from Scenario 3.
- In Table A1, we summarize the results for $\hat{\boldsymbol{\beta}}$ under the null of no intervention effect. In addition to bias (times 1000) and ratio between mean estimated standard error and empirical standard deviation, Table A1 gives the proportion of times the null $H_0 : \beta_1 = 0$ in model (1) of the main paper was rejected ($\hat{\alpha}_\beta$) and the proportion of times the null $H_0 : \gamma = 0$ in the model logit $\tilde{p}_Q(\boldsymbol{\beta}, \gamma; \boldsymbol{z}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_2^T \boldsymbol{z} + \gamma Q$, where Q is a group indicator that equals one for the intervention group and zero for the control, was rejected. When the number of centers was small (three in each arm), the standard error was underestimated. The type I error was generally close to the desired 0.05 level and was between 0.04 and 0.06 as expected given the number of simulation iterations (2000).
- Tables A2, A3 and A4 are extended versions of Tables 1, 2 and 3 in the main paper, presenting the results of more simulation scenarios.
- Table A5 presents the mean estimated success probability of the estimated optimal intervention package, the mean true success probability of the estimated optimal intervention and the mean estimated success probabili ty of the true optimal intervention. In all simulation scenarios, All these mean probabilities were very close to the desired 90% ($\tilde{p}$).
- Tables A6 and A7 are the analogues of Tables 2 and 3 in the main paper and Tables A3 and A4 in this document, when changing the cost function as follows. For the second component of the intervention, the unit cost in Table A6 is $c_2 = 6$ instead of $c_2 = 8$, while keeping $c_1 = 1$. In the case of $c_2 = 6$, β_{11}/c_1 and β_{12}/c_1 are closer to each other, and the finite sample bias is larger than when $c_2 = 8$. However, as can be seen from Table A6, the bias was greatly reduced when the per-center sample size or the number of centers were larger.
- Tables A8 and A9 are the analogues of Tables 2 and 3 in the main paper and Tables A3 and A4 in this document, when changing the cost function as follows. Tables A8 and A9 consider a non-linear cost function for the intervention. As an example, we considered $C(\boldsymbol{x}) = x_1 + 5x_2 + 3\exp(-0.5x_2)$. That is, the cost is linear in the first components, and has a diminishing marginal cost for the second component. In this case, as can be seen from Tables A8 and A9 , the optimal intervention is not necessarily on the boundary. The bias in the estimation of the optimal intervention was not too large, and was further reduced as the sample size and number of centers increased. The empirical coverage of the confidence set $CS(\boldsymbol{x}^{opt})$ was slightly lower than the expected 95%.

5.2.1 Scenario 3 simulation results

As written in the main text, data were simulated 2000 times mimicking the three stages of the BetterBirth study. True coefficient values were taken from the last column of Table 4 in the main text. That is, $\beta_{11} = 1.03$ for launch duration, $\beta_{12} = 0.026$ for each coaching visit, $\beta_0 = -2.3$ as the intercept, and $\beta_2 = 0.66$ as the coefficient for mean birth volume. The % relative bias of the two estimated coefficient effects $\hat{\boldsymbol{\beta}}$ in each stage were 18.4% and -77.3% after stage 1, 2.3% and -50.9% after stage 2, and -0.2% and -7.4% for the final estimators. The observed relative biases in the final stages were comparable with the relative bias reported in Table A.2. The first and second stage bias were larger in absolute value.

The optimal intervention package for a center with average birth volume ($z = 1.75$) was a launch duration of 2.78 days and 1 coaching visit $\boldsymbol{x}^{opt} = (2.78, 1)$. In the 2000 simulation iterations, the mean stage 2 recommended intervention was $(2.21, 2.93)$, the mean stage 3 recommended intervention was $(2.64, 1.37)$, and the mean final estimated optimal intervention after all three stages was $(2.78, 1)$, practically the same as the true optimal intervention. If only combinations of discrete values within $\mathcal{X}$ are allowed, that is $1, ..., 40$ for coaching visits and $1, 1.5, 2, 2.5, ..., 5$ for duration of intervention launch, then the optimal intervention is launch duration of three days with one coaching visit, $\boldsymbol{x}^{opt}_{disc} = (3, 1)$. In the 2000 simulation iterations, the mean stage 2 recommended discrete intervention was $(2.31, 3.10)$, the mean stage 3 recommended discrete intervention was $(2.84, 1.60)$, and the mean final estimated optimal discrete intervention after all three stages was $(3.001, 1.01)$, practically the same as the true optimal intervention. The overall conclusion from simulation scenario 3 was the LAGO would perform well under settings similar to the BetterBirth study.

Table A1: Simulation study results under the null $\beta_1 = 0$. Unit costs were $c_1 = 1$ and $c_2 = 8$. Results presented are bias times 1000, mean estimated standard error to empirical standard deviation ratio (times 100) and type I error using tests for $H_0 : \boldsymbol{\beta}_1 = 0$ (α_β) and for $H_0 : \gamma = 0$ (α_γ).

n_{1j}	n_{2j}	J	$\hat{\beta}_{11}$ Bias (×1000)	$\hat{\beta}_{11}$ SE/EMP.SD	$\hat{\beta}_{12}$ Bias (×1000)	$\hat{\beta}_{12}$ SE/EMP.SD	$\hat{\alpha}_\beta$	$\hat{\alpha}_\gamma$
Scenario 1 ($J_1 = J_2 = J$)								
50	100	6	-10.4	89.2	-16.5	93.6	3.9	4.9
		10	-6.2	98.4	-13.0	101.0	4.6	5.1
		20	-5.0	100.0	-6.8	102.9	4.5	5.4
	200	6	-6.4	74.2	-18.5	87.3	5.3	4.9
		10	-6.2	98.3	-11.5	96.1	4.2	5.5
		20	-6.9	98.6	-6.1	98.5	5.1	4.7
	500	6	-13.3	79.7	-13.4	89.5	5.5	5.3
		10	-6.2	92.0	-9.9	91.0	4.8	5.1
		20	-4.6	95.9	-5.1	94.5	4.9	5.1
	1000	6	-13.3	80.5	-9.4	85.7	4.8	5.0
		10	-6.4	92.1	-6.7	89.3	4.2	4.0
		20	-3.2	91.0	-3.4	94.0	5.0	5.6
100	100	6	-19.0	30.3	3.2	14.7	4.5	5.9
		10	-10.2	96.3	-4.0	95.4	5.6	5.4
		20	-3.9	100.7	-5.1	101.7	4.7	4.2
	200	6	-7.2	87.7	-7.5	92.7	5.3	4.7
		10	-6.2	93.5	-9.1	93.3	5.0	5.0
		20	-4.3	101.2	-5.8	97.6	5.7	5.8
	500	6	-9.3	84.8	-10.2	92.0	4.0	4.7
		10	-6.2	96.4	-6.0	93.1	4.9	5.0
		20	-4.6	96.6	-4.0	98.9	5.0	4.5
	1000	6	-7.0	86.9	-7.4	88.6	5.2	5.0
		10	-4.5	93.8	-6.0	95.6	4.2	5.4
		20	-2.9	94.9	-4.3	96.4	5.4	5.1
200	100	6	-6.9	89.0	-7.9	91.4	4.5	4.6
		10	-4.9	98.2	-4.5	96.9	4.9	4.9
		20	-2.1	101.4	-3.0	98.6	5.2	4.8
	200	6	-9.8	92.1	-4.6	93.6	4.3	4.8
		10	-4.0	103.1	-4.8	99.5	3.9	4.9
		20	-2.9	104.3	-3.8	101.7	4.6	4.4
	500	6	-10.0	47.1	-6.2	46.6	5.0	5.1
		10	-4.1	90.9	-5.6	98.1	4.9	4.4
		20	-3.1	101.7	-3.0	101.7	4.2	5.0
	1000	6	-3.9	91.1	-6.5	85.4	5.1	4.8
		10	-6.0	93.6	-3.3	92.7	4.2	4.7
		20	-2.5	103.7	-3.2	100.2	4.3	4.9
Scenario 2a ($J_1 = 6,\ J_2 = 12$)								
50	200		-3.1	94.6	-11.0	93.8	4.9	4.7
Scenario 2b ($J_1 = 10,\ J_2 = 20$)								
50	200		-7.4	94.7	-9.1	93.2	5.1	5.4

Bias, mean bias $\hat{\beta} - \beta^\star$; SE, mean estimated standard error;
EMP.SD, empirical standard deviation; CP95, empirical coverage rate of 95% confidence intervals.

Table A2: Extended version of Table 1 in the main paper. Simulation study results for individual package component effects. Unit costs were $c_1 = 1$ and $c_2 = 8$. Results include for each individual component estimator bias (times 1000), ratio between mean estimated standard error and empirical standard deviation (times 100) and coverage rate of 95% confidence interval.

$\exp(\boldsymbol{\beta}^\star)$	n_{1j}	n_{2j}	J		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$	
				Bias ($\times$1000)	SE/EMP.SD	CP95	Bias ($\times$1000)	SE/EMP.SD	CP95
Scenario 1 ($J_1 = J_2 = J$)									
$(1, 1.2)$	50	100	6	-3.0	91.1	95.4	-17.3	74.6	95.0
			10	0.0	103.0	96.5	-12.1	92.2	95.2
			20	-2.8	99.7	95.3	-6.2	96.6	95.4
		200	6	-4.9	93.7	95.8	-14.3	76.3	94.7
			10	-1.7	94.6	94.4	-10.2	87.2	94.4
			20	-5.2	96.8	94.3	-3.1	94.5	94.8
		500	6	-10.0	84.9	96.5	-11.2	61.3	95.2
			10	-7.1	96.5	95.3	-4.9	86.9	94.6
			20	-4.4	97.5	95.2	-1.6	98.8	95.8
		1000	6	-3.9	76.7	95.6	-8.9	76.5	95.0
			10	-7.0	92.2	94.4	-3.2	80.2	94.3
			20	-3.5	94.0	94.7	-1.2	91.3	95.0
	100	100	6	0.6	94.0	94.2	-13.4	88.5	95.5
			10	1.0	98.8	94.8	-9.0	92.8	94.8
			20	-4.0	97.4	94.1	-3.2	98.1	95.0
		200	6	-6.8	90.4	96.1	-10.3	85.0	95.1
			10	-5.0	98.8	95.4	-5.8	93.4	96.0
			20	-1.7	99.0	94.8	-3.1	99.7	95.5
		500	6	-8.6	82.1	94.1	-9.0	77.5	94.4
			10	-3.6	94.5	94.9	-4.9	84.2	94.8
			20	-3.7	94.9	94.0	-1.8	95.1	94.0
		1000	6	-3.6	86.6	95.2	-8.0	77.6	94.2
			10	-4.3	93.2	95.3	-3.5	84.3	95.8
			20	-3.6	98.3	94.7	-0.8	97.6	95.5
	200	100	6	2.8	93.4	95.0	-9.7	92.8	95.5
			10	-2.4	100.8	96.0	-3.2	101.0	95.9
			20	-0.8	99.3	94.7	-1.9	101.2	95.7
		200	6	-2.5	90.4	94.7	-9.7	87.8	95.0
			10	-1.5	100.1	95.2	-4.2	98.8	95.2
			20	-0.9	98.5	94.7	-1.2	98.4	94.7
		500	6	-2.0	92.5	94.7	-6.8	86.3	94.6
			10	-4.3	96.7	94.8	-2.4	94.5	95.8
			20	-2.4	98.6	95.0	-1.5	98.0	94.6
		1000	6	-4.4	86.6	95.1	-5.3	77.8	94.5
			10	-0.9	95.0	95.0	-2.8	91.4	95.3
			20	-3.8	93.9	94.0	-0.1	99.6	94.7
$(1, 1.5)$	50	100	6	2.2	85.6	94.8	-15.1	82.1	94.3
			10	3.8	94.4	95.1	-7.8	88.6	94.0
			20	2.5	96.6	94.8	-5.3	95.8	94.6
		200	6	-7.3	86.5	94.8	-11.5	72.8	94.4
			10	-4.2	96.6	95.4	-4.8	89.9	94.5
			20	-1.0	98.4	95.9	-3.3	94.2	94.6
		500	6	-3.6	85.9	95.3	-7.4	71.7	94.0
			10	-1.2	90.9	94.3	-4.4	86.0	94.4
			20	-2.5	95.7	94.5	-1.3	93.1	94.0
		1000	6	-2.8	76.0	95.5	-6.4	77.3	94.5
			10	-1.8	92.2	95.5	-2.5	82.1	94.0
			20	-3.3	92.2	94.0	-1.0	92.5	94.2
	100	100	6	9.1	88.5	94.7	-15.0	79.2	92.9
			10	3.0	96.2	95.0	-5.7	89.3	94.0
			20	2.8	96.9	94.8	-4.1	97.2	95.1
		200	6	4.2	92.0	95.1	-12.7	86.9	94.2
			10	-0.1	96.6	95.0	-4.6	87.4	94.5
			20	0.6	97.1	95.7	-2.1	93.9	95.2
		500	6	-4.4	80.8	94.7	-6.2	74.9	94.0
			10	-2.5	93.7	94.7	-3.2	86.1	94.8
			20	-0.8	96.9	94.4	-2.1	96.0	95.4
		1000	6	-3.8	91.4	95.6	-3.8	74.8	94.6
			10	-1.7	95.2	94.5	-2.2	84.4	94.6
			20	-1.0	97.1	94.9	-1.4	94.3	95.0
	200	100	6	1.7	91.9	95.7	-8.7	81.6	94.7
			10	2.7	98.3	95.1	-4.8	93.2	95.1
			20	1.2	97.8	95.7	-1.5	99.0	95.7
		200	6	0.5	91.1	95.6	-6.8	83.7	94.8
			10	1.8	92.9	94.8	-4.0	91.0	95.0
			20	2.8	95.5	94.3	-2.7	93.3	93.7
		500	6	1.0	89.7	94.5	-5.5	77.4	94.3
			10	-0.7	96.0	95.2	-2.3	90.5	94.4
			20	-0.6	95.5	94.7	-1.1	93.7	94.4
		1000	6	0.1	90.4	95.0	-5.3	73.9	94.7
			10	-0.6	94.5	94.6	-2.3	90.7	94.2
			20	-1.6	95.6	94.5	-1.3	95.0	94.2

$\exp(\boldsymbol{\beta}^\star)$	n_{1j}	n_{2j}	J	$\hat{\beta}_{11}$ Bias (×1000)	$\hat{\beta}_{11}$ SE/EMP.SD	$\hat{\beta}_{11}$ CP95	$\hat{\beta}_{12}$ Bias (×1000)	$\hat{\beta}_{12}$ SE/EMP.SD	$\hat{\beta}_{12}$ CP95
$(1,2)$	50	100	6	-20.3	87.5	95.3	3.0	88.6	94.1
			10	-6.5	100.6	95.6	-0.7	96.2	95.0
			20	-1.8	97.9	94.5	-2.6	93.7	94.1
		200	6	-10.1	89.3	95.0	-0.6	82.9	95.2
			10	-10.5	99.3	95.3	-3.2	93.4	95.0
			20	-5.1	99.0	95.0	-1.8	94.5	95.6
		500	6	-13.9	88.9	94.6	-3.0	82.5	94.8
			10	-9.6	94.6	95.2	-1.1	91.8	93.8
			20	-4.8	96.2	94.8	-1.1	91.4	94.8
		1000	6	-9.8	83.4	94.9	-3.2	79.9	94.1
			10	-8.4	94.7	95.5	-0.6	88.8	94.8
			20	-2.7	96.4	94.8	-1.1	91.2	94.8
	100	100	6	-8.2	91.1	94.8	-2.6	83.4	95.2
			10	-6.2	94.3	95.0	-1.1	93.2	94.8
			20	-2.1	96.3	94.0	-2.1	92.8	94.2
		200	6	-6.0	88.6	95.4	-0.4	90.0	95.2
			10	-4.2	97.6	95.3	-2.3	95.0	94.9
			20	-1.7	100.4	95.6	-0.8	94.8	95.8
		500	6	-2.1	89.8	95.0	-4.0	90.5	94.8
			10	-3.2	99.0	95.5	-2.7	93.0	95.2
			20	-1.1	100.7	95.2	-1.1	93.5	95.2
		1000	6	-7.4	93.1	95.9	0.2	80.8	95.9
			10	-3.3	94.6	94.5	-2.3	87.5	94.5
			20	-2.0	102.2	96.0	-1.5	92.9	94.2
	200	100	6	-2.7	90.3	95.4	0.3	92.9	95.5
			10	-0.6	94.6	95.0	-1.8	94.5	95.1
			20	-0.1	95.3	95.2	-0.6	92.5	94.4
		200	6	-9.3	85.3	94.9	0.0	89.0	95.0
			10	-2.9	94.0	94.2	-1.7	93.7	95.5
			20	0.1	96.0	95.0	-2.0	91.9	93.8
		500	6	-2.0	90.7	94.8	-3.5	87.3	95.3
			10	-2.8	94.5	94.7	-1.9	90.8	95.0
			20	0.5	98.8	94.9	-0.6	91.1	94.3
		1000	6	-2.3	87.2	95.3	-3.8	81.5	94.7
			10	-0.4	96.7	95.0	-1.6	89.1	94.3
			20	0.7	99.2	95.2	-1.3	91.6	94.2
$(1.2,1.5)$	50	100	6	-4.2	96.5	95.1	-7.7	84.1	94.0
			10	-5.0	98.8	94.9	-5.0	92.2	95.2
			20	-2.5	101.3	95.2	-1.3	102.7	95.6
		200	6	-3.2	95.0	94.9	-10.4	81.0	95.4
			10	-8.0	92.7	94.2	-3.9	91.9	95.2
			20	-3.8	102.2	95.5	-0.9	99.7	95.2
		500	6	-9.9	87.2	95.2	-4.1	82.0	95.5
			10	-3.7	95.3	95.3	-2.2	86.7	95.2
			20	-2.5	96.6	94.7	-0.8	98.3	95.5
		1000	6	-10.6	84.3	94.3	-3.6	75.9	95.3
			10	-6.4	88.2	94.0	-1.8	85.2	94.0
			20	-2.0	96.9	94.9	-0.1	95.8	94.8
	100	100	6	-3.1	92.9	94.7	-6.0	86.2	95.5
			10	5.1	101.9	95.7	-5.5	100.9	95.4
			20	3.9	101.1	95.5	-2.2	101.6	95.0
		200	6	-5.9	91.4	94.6	-3.2	83.6	95.5
			10	-2.9	99.5	95.4	-2.4	94.9	95.3
			20	-0.8	98.4	95.0	-1.1	97.5	94.5
		500	6	-6.5	91.1	94.6	-2.9	88.3	94.8
			10	-0.3	99.7	95.2	-2.8	94.7	94.8
			20	-2.4	102.5	95.8	-0.4	102.5	95.8
		1000	6	-4.2	89.7	95.0	-3.0	80.5	95.8
			10	-2.7	94.6	94.2	-2.0	90.2	95.2
			20	-1.6	99.9	95.4	0.1	98.1	95.3
	200	100	6	-3.2	91.6	95.0	-2.5	91.8	95.0
			10	1.1	102.4	96.0	-2.0	100.9	95.5
			20	1.0	100.7	95.3	0.0	98.7	95.2
		200	6	2.4	95.7	95.3	-5.5	93.0	95.0
			10	-0.2	98.0	94.9	-1.6	98.0	96.0
			20	1.6	98.5	95.1	-0.2	101.7	95.5
		500	6	-0.5	93.3	94.8	-4.6	93.5	94.8
			10	-1.1	95.0	94.3	-1.1	97.1	94.9
			20	-1.2	97.6	94.4	0.2	98.1	94.8
		1000	6	-2.3	91.3	95.1	-2.3	78.6	94.8
			10	-1.8	97.6	95.2	-0.7	94.7	95.5
			20	-1.2	98.6	94.5	0.3	97.2	94.2

$\exp(\boldsymbol{\beta}^\star)$	n_{1j}	n_{2j}	J	$\hat\beta_{11}$ Bias ($\times$1000)	SE/EMP.SD	CP95	$\hat\beta_{12}$ Bias ($\times$1000)	SE/EMP.SD	CP95
$(1.2, 2)$	50	100	6	-29.2	91.6	95.4	5.1	86.0	96.0
			10	-13.6	101.4	95.8	1.6	102.2	96.0
			20	-6.5	99.6	95.2	-0.9	101.4	94.8
		200	6	-21.6	89.9	95.1	4.7	89.7	95.1
			10	-16.7	94.9	95.5	0.9	97.6	96.0
			20	-4.8	100.0	95.0	-1.4	101.4	96.2
		500	6	-19.0	80.1	95.1	-1.2	77.7	95.5
			10	-12.4	97.1	95.5	1.4	99.5	96.5
			20	-6.7	94.4	94.6	-0.8	101.8	96.0
		1000	6	-13.2	83.3	95.2	-1.0	82.0	94.4
			10	-9.2	92.3	94.9	0.5	94.2	95.5
			20	-4.9	96.4	95.2	0.1	96.4	94.8
	100		6	-13.8	94.5	95.8	-0.8	94.1	95.2
			10	-3.8	98.2	94.8	-0.1	102.7	95.2
			20	-6.8	100.3	95.2	1.5	102.7	95.5
		200	6	-13.0	84.6	95.2	2.1	95.8	95.9
			10	-8.4	96.4	94.7	0.2	99.6	95.5
			20	-6.4	98.0	94.6	0.7	104.8	95.9
		500	6	-13.7	89.2	95.4	0.2	97.8	95.2
			10	-3.6	96.5	95.5	-2.1	98.3	95.6
			20	-5.4	97.5	95.5	0.1	100.5	95.3
		1000	6	-7.2	87.8	94.9	-0.3	83.4	94.6
			20	-4.1	95.4	94.5	-1.1	96.6	95.3
			10	-4.3	96.0	94.8	0.4	101.1	95.0
	200	100	6	-11.8	87.6	96.2	3.5	93.8	95.0
			10	-3.1	96.8	95.5	0.3	106.4	96.5
			20	-1.1	100.4	95.2	-0.1	103.6	95.8
		200	6	-6.9	93.2	95.8	-0.1	96.2	95.8
			10	-1.6	98.7	95.4	-1.5	102.1	95.2
			20	-0.5	98.7	95.2	-0.5	105.5	95.7
		500	6	-9.3	87.9	95.5	1.1	85.5	96.5
			10	-3.8	96.4	95.2	-1.0	101.6	95.5
			20	-2.6	100.4	95.1	1.0	103.3	95.2
		1000	6	-8.5	83.5	95.0	-0.6	94.5	95.7
			10	-6.2	97.3	95.7	-0.3	101.3	95.8
			20	-1.6	98.2	95.3	-0.1	102.0	95.7
Scenario 2a ($J_1 = 6,\ J_2 = 12$)									
$(1.0, 1.2)$	50	200		-6.2	92.1	94.8	-8.7	79.5	94.0
$(1.0, 1.5)$				-6.2	97.2	95.3	-5.0	82.0	94.9
$(1.0, 2.0)$				-7.0	93.3	95.5	-0.7	89.1	94.7
$(1.2, 1.5)$				-6.9	96.4	95.5	-2.0	91.0	94.8
$(1.2, 2.0)$				-13.5	95.6	95.9	5.0	94.7	95.5
Scenario 2b ($J_1 = 10,\ J_2 = 20$)									
$(1.0, 1.2)$	50	200	5.0	-1.7	96.4	95.6	-6.0	85.1	94.6
$(1.0, 1.5)$				-4.7	96.3	94.8	-2.4	90.7	94.2
$(1.0, 2.0)$				-4.5	95.7	94.5	-2.4	94.3	94.2
$(1.2, 1.5)$				-5.6	96.9	94.6	-2.6	95.5	95.5
$(1.2, 2.0)$				-11.2	93.4	94.7	1.1	100.1	95.2

Table A3: Extended version of Table 2 in the main paper. Simulation results for estimated optimal package. Unit costs were $c_1 = 1$ and $c_2 = 8$. Results include bias for optimal individual components (times 100) and empirical mean squared error (times 100) of the second-stage recommended intervention and the final estimated optimal package.

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	Stage 1 Bias$_1$ (×100)	Stage 1 Bias$_2$ (×100)	Stage 1 RMSE (×100)	Stage 2 Bias$_1$ (×100)	Stage 2 Bias$_2$ (×100)	Stage 2 RMSE (×100)
$(0, 3.2)$	$(2, 4.5)$	50	100	6	75.3	-47.0	170.8	57.3	-14.2	116.2
				10	63.0	-20.2	131.9	47.1	-8.9	100.7
				20	52.8	-10.0	110.6	34.5	-4.7	85.0
			200	6	76.4	-46.3	169.6	52.2	-12.9	109.5
				10	66.9	-18.2	134.6	40.3	-6.4	92.3
				20	53.4	-9.5	111.1	23.8	-3.1	70.5
			500	6	75.9	-47.9	170.5	42.5	-7.9	97.4
				10	67.9	-19.6	134.7	29.6	-4.9	78.8
				20	52.6	-11.5	110.5	16.5	-2.1	58.5
			1000	6	79.2	-44.4	173.4	34.9	-6.1	87.6
				10	68.8	-18.3	136.1	22.7	-3.6	68.8
				20	51.7	-9.9	109.9	9.1	-0.8	43.5
		100	100	6	70.0	-30.7	150.5	55.6	-10.8	111.9
				10	54.8	-14.7	115.9	39.4	-6.7	91.8
				20	35.0	-5.8	89.0	24.0	-2.5	71.0
			200	6	71.3	-29.2	150.5	49.5	-10.0	104.2
				10	58.5	-11.5	119.5	35.3	-5.2	86.2
				20	34.0	-6.4	87.5	16.7	-2.0	59.2
			500	6	70.8	-32.4	151.5	38.0	-6.6	90.9
				10	63.0	-12.4	123.8	26.1	-3.2	73.8
				20	38.9	-7.5	93.0	10.6	-0.9	47.0
			1000	6	70.1	-30.5	149.7	30.8	-6.1	81.1
				10	57.9	-12.7	119.1	19.5	-2.3	63.7
				20	36.1	-5.6	90.0	5.2	0.1	33.1
		200	100	6	59.4	-22.2	130.6	43.3	-8.5	97.9
				10	41.8	-7.9	98.2	30.0	-4.0	80.0
				20	21.8	-2.1	69.6	15.1	-1.5	56.7
			200	6	61.8	-21.9	134.2	35.8	-6.7	88.8
				10	44.2	-8.5	102.1	28.2	-3.7	76.9
				20	22.6	-2.5	70.7	11.2	-0.4	48.8
			500	6	63.1	-20.3	133.1	33.8	-4.6	84.5
				10	40.6	-10.5	97.5	18.6	-2.0	62.4
				20	21.6	-3.7	69.3	7.0	-0.5	38.4
			1000	6	62.6	-22.7	134.0	28.6	-3.8	77.6
				10	45.9	-8.6	102.1	14.6	-1.4	55.1
				20	22.2	-3.2	69.8	3.8	0.2	28.4
$(1.2, 1.5)$	$(2, 4.5)$	50	100	6	-68.5	-88.2	208.6	-37.0	-7.0	105.7
				10	-48.4	-36.4	142.8	-26.6	2.1	81.6
				20	-30.0	-9.9	94.5	-9.5	2.7	51.6
			200	6	-69.4	-94.7	213.3	-32.0	-1.2	95.9
				10	-47.1	-33.6	139.8	-22.1	4.3	73.7
				20	-27.9	-8.3	90.4	-5.7	3.3	39.8
			500	6	-69.8	-90.8	210.0	-24.3	0.0	82.7
				10	-49.8	-36.3	143.8	-9.8	2.7	51.2
				20	-30.7	-9.8	94.8	-2.7	2.1	27.8
			1000	6	-65.5	-95.0	212.7	-17.1	2.2	68.4
				10	-52.8	-35.4	144.2	-6.2	3.8	41.5
				20	-27.5	-9.6	91.5	-0.5	1.1	15.8
		100	100	6	-54.4	-57.1	174.1	-31.8	-4.9	97.1
				10	-32.2	-15.2	104.1	-14.5	2.2	63.4
				20	-14.9	-3.1	68.6	-3.6	1.2	35.9
			200	6	-55.9	-57.5	171.9	-25.8	-0.8	84.6
				10	-35.2	-14.8	107.4	-11.9	3.2	56.3
				20	-15.8	-2.4	69.5	-3.1	2.2	31.0
			500	6	-52.5	-57.8	171.7	-18.7	2.6	69.1
				10	-33.4	-17.8	109.2	-5.5	3.0	39.8
				20	-16.6	-2.5	70.9	-0.7	1.7	18.1
			1000	6	-60.3	-55.5	173.4	-13.3	1.9	59.7
				10	-37.5	-18.8	115.0	-3.4	3.0	31.9
				20	-14.6	-3.4	68.1	-0.3	0.7	13.2
		200	100	6	-44.6	-35.8	142.4	-23.2	-1.3	79.3
				10	-21.7	-4.9	81.3	-8.4	1.9	50.5
				20	-5.1	0.1	45.5	-1.3	0.3	27.3
			200	6	-46.3	-32.8	142.6	-16.2	2.2	67.5
				10	-21.9	-7.8	85.3	-7.6	2.1	46.4
				20	-5.3	-0.4	46.3	-0.6	-0.1	19.9
			500	6	-40.5	-34.3	139.4	-12.8	3.4	57.8
				10	-24.1	-6.8	86.4	-4.2	1.9	34.9
				20	-6.4	-0.4	48.5	-0.2	0.6	13.9
			1000	6	-41.5	-31.0	136.5	-8.3	2.1	48.0
				10	-19.8	-5.8	78.6	-1.7	1.7	24.0
				20	-4.9	0.0	44.0	-0.1	0.4	10.3

Table A.3: continued

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	Stage 1			Stage 2		
					Bias_1 ($\times$100)	Bias_2 ($\times$100)	RMSE ($\times$100)	Bias_1 ($\times$100)	Bias_2 ($\times$100)	RMSE ($\times$100)
$(1.2, 2)$	$(2, 2.6)$	50	100	6	-83.2	-27.6	166.0	-68.2	3.2	125.3
				10	-67.7	-6.9	132.4	-53.9	4.7	107.6
				20	-50.2	-0.5	106.3	-33.1	4.5	84.0
			200	6	-80.4	-30.9	164.3	-59.0	3.1	114.9
				10	-67.6	-7.0	133.0	-43.3	5.1	96.4
				20	-45.6	-0.3	101.6	-23.2	3.7	70.3
			500	6	-79.7	-27.2	163.2	-48.5	4.5	103.8
				10	-69.2	-6.9	132.9	-30.9	4.2	81.2
				20	-51.4	0.5	107.1	-14.9	3.3	56.6
			1000	6	-82.0	-35.2	167.3	-38.5	4.1	92.4
				10	-64.7	-7.5	131.1	-22.9	3.5	69.9
				20	-50.4	0.3	106.6	-7.1	2.0	39.3
		100	100	6	-71.6	-16.8	146.3	-54.9	3.8	110.8
				10	-52.1	-2.5	111.7	-38.6	2.6	90.9
				20	-35.8	1.7	88.2	-23.2	3.3	70.3
			200	6	-71.1	-18.1	144.5	-51.6	2.7	106.6
				10	-55.5	0.1	114.3	-34.6	3.9	86.0
				20	-37.4	1.9	90.2	-17.4	3.0	61.0
			500	6	-72.6	-12.5	145.2	-40.3	4.1	93.1
				10	-53.4	-1.7	112.7	-24.6	3.4	72.3
				20	-35.0	1.7	87.5	-8.8	2.3	43.6
			1000	6	-75.3	-18.6	145.3	-34.3	3.0	86.1
				10	-57.1	-2.5	116.2	-16.9	2.8	60.2
				20	-35.2	1.9	87.4	-5.5	1.6	34.5
		200	100	6	-62.8	-9.6	130.1	-46.6	2.1	100.8
				10	-43.4	1.0	98.5	-31.3	2.8	81.8
				20	-19.2	1.4	64.6	-10.6	1.5	48.1
			200	6	-60.4	-11.4	127.8	-40.2	3.1	93.3
				10	-42.1	-0.0	96.8	-24.1	2.8	71.8
				20	-18.2	0.7	62.9	-7.8	1.2	41.4
			500	6	-61.0	-9.4	126.9	-32.9	3.3	84.1
				10	-40.5	0.1	95.3	-19.1	3.0	63.8
				20	-17.8	1.3	62.4	-4.4	0.8	31.0
			1000	6	-63.2	-10.8	129.3	-26.2	3.6	75.3
				10	-40.5	1.1	95.1	-12.0	2.8	50.7
				20	-18.8	1.0	63.8	-1.7	0.7	19.7

Table A4: Extended version of Table 3 in the main paper. Simulation results for estimated optimal package and coverage of confidence bands for success probabilities. Scenario 1 with unit costs $c_1 = 1$ and $c_2 = 8$. Results include empirical 2.5% and 97.5% quantiles of true success rate of second-stage recommended intervention and final estimated optimal intervention, coverage rate of 95% confidence set for optimal intervention package (SetCP95), and its relative size (SetPerc%). Results also include the coverage rate of 95% confidence bands for outcome probability under all values on a grid of $\mathcal{X}$ (BandsCP95) and the ratio between the mean cost of the estimated optimal intervention and the cost of the optimal true intervention (RelCost).

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	PrOpt1 (Q2.5,Q97.5)	PrOpt2 (Q2.5,Q97.5)	SetCP95	SetPerc%	BandsCP95	RelCost
Scenario 1 ($J_1 = J_2 = 20$)										
$(1, 2)$	$(0, 3.2)$	50	100	6	(50.0,97.0)	(82.5,93.5)	92.9	16.4	98.0	97.8
				10	(77.1,96.0)	(85.8,92.3)	94.8	11.5	97.3	99.1
				20	(83.6,93.8)	(87.2,91.8)	94.0	7.6	97.0	99.9
			200	6	(50.0,97.0)	(83.5,92.8)	93.2	13.8	96.6	98.0
				10	(77.5,95.7)	(86.6,91.8)	95.2	8.8	97.8	99.6
				20	(83.7,93.8)	(88.0,91.4)	95.8	5.8	97.5	100.0
			500	6	(50.0,97.0)	(85.0,92.3)	94.1	10.2	97.4	99.2
				10	(78.1,95.9)	(87.1,91.4)	94.5	6.4	96.8	99.6
				20	(83.5,93.7)	(88.2,91.1)	95.0	4.0	97.2	100.0
			1000	6	(50.0,97.0)	(85.6,91.8)	94.6	8.4	96.5	99.5
				10	(78.5,96.3)	(87.6,91.1)	95.0	5.0	97.5	99.8
				20	(83.6,93.7)	(88.7,91.0)	95.2	3.1	97.0	100.1
		100		6	(50.0,97.0)	(84.5,93.0)	93.8	14.1	97.0	98.8
				10	(81.7,94.3)	(86.4,92.1)	94.4	9.5	97.8	99.4
				20	(85.2,93.1)	(87.8,91.6)	94.8	6.3	96.5	100.2
			200	6	(55.1,97.0)	(85.1,92.2)	94.3	11.7	97.4	98.8
				10	(81.3,94.6)	(87.1,91.8)	94.3	7.7	97.0	99.8
				20	(85.5,92.9)	(88.1,91.3)	95.2	5.1	97.5	100.0
			500	6	(50.0,97.0)	(86.1,92.0)	94.2	9.0	96.8	99.4
				10	(81.7,94.8)	(87.7,91.4)	95.4	5.8	96.8	100.0
				20	(85.6,92.8)	(88.8,91.0)	95.3	3.7	97.4	100.1
			1000	6	(50.0,96.8)	(86.7,91.4)	95.6	6.8	97.7	99.3
				10	(82.0,94.6)	(88.0,91.1)	94.6	4.6	96.8	100.0
				20	(85.5,93.0)	(89.1,90.9)	94.5	2.9	97.9	100.2
		200	100	6	(68.6,95.2)	(85.1,92.5)	94.6	11.8	97.4	99.0
				10	(84.8,93.3)	(87.1,92.0)	95.2	8.1	97.8	99.9
				20	(87.0,92.4)	(88.2,91.5)	94.5	5.3	97.0	100.1
			200	6	(62.9,95.2)	(86.0,92.2)	94.6	9.6	97.0	99.3
				10	(84.2,93.3)	(87.6,91.6)	94.6	6.6	97.2	100.0
				20	(87.0,92.2)	(88.4,91.4)	94.2	4.4	96.9	100.3
			500	6	(70.6,95.1)	(87.2,91.8)	95.4	7.8	97.2	99.9
				10	(83.8,93.1)	(88.2,91.3)	95.1	5.0	97.0	100.1
				20	(87.0,92.2)	(88.9,91.0)	94.4	3.4	96.8	100.1
			1000	6	(69.8,95.1)	(87.3,91.5)	95.4	6.2	98.0	99.9
				10	(84.4,93.1)	(88.5,91.2)	94.4	4.0	97.2	100.1
				20	(87.0,92.1)	(89.2,90.8)	94.6	2.6	97.0	100.2

Table A.4: continued

$e^{\boldsymbol{\beta}^{\star}}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	PrOpt1 (Q2.5,Q97.5)	PrOpt2 (Q2.5,Q97.5)	SetCP95	SetPerc%	BandsCP95	RelCost
Scenario 1 ($J_1 = J_2 = 20$)										
$(1.2, 1.5)$	$(2, 4.5)$	50	100	6	(56.9,91.6)	(81.5,91.6)	94.0	24.4	95.8	97.6
				10	(68.9,91.6)	(85.2,91.6)	94.2	18.4	96.2	99.7
				20	(81.1,91.6)	(87.3,91.6)	94.8	13.3	96.0	100.3
			200	6	(56.5,91.6)	(83.3,91.6)	95.0	21.4	96.2	98.9
				10	(70.7,91.6)	(85.9,91.6)	95.5	15.1	96.6	100.3
				20	(82.6,91.6)	(88.1,91.6)	94.5	10.6	94.8	100.5
			500	6	(56.7,91.6)	(84.7,91.6)	94.8	17.4	96.8	99.4
				10	(68.4,91.6)	(87.1,91.6)	95.4	11.9	95.6	100.3
				20	(81.9,91.6)	(88.8,91.3)	95.1	7.6	95.9	100.4
			1000	6	(56.5,91.6)	(85.7,91.6)	95.1	14.9	95.9	100.0
				10	(71.2,91.6)	(88.0,91.6)	93.7	9.1	94.6	100.6
				20	(81.4,91.6)	(89.2,91.0)	94.9	5.6	94.2	100.2
		100	100	6	(58.6,91.6)	(83.0,91.6)	93.8	21.9	96.7	98.1
				10	(79.4,91.6)	(86.5,91.6)	94.5	16.8	96.3	100.1
				20	(84.7,91.6)	(87.9,91.6)	94.8	12.3	95.4	100.2
			200	6	(59.0,91.6)	(84.5,91.6)	94.6	19.1	96.4	99.1
				10	(79.4,91.6)	(86.9,91.6)	94.6	13.9	96.2	100.3
				20	(84.4,91.6)	(88.4,91.5)	94.9	9.9	94.9	100.4
			500	6	(59.0,91.6)	(85.9,91.6)	94.4	15.4	95.5	100.0
				10	(78.6,91.6)	(88.0,91.6)	95.4	10.7	95.8	100.5
				20	(84.0,91.6)	(89.0,91.1)	95.3	7.1	95.4	100.3
			1000	6	(58.9,91.6)	(86.3,91.6)	94.7	12.8	96.2	100.1
				10	(77.4,91.6)	(88.4,91.6)	94.6	8.3	95.0	100.5
				20	(84.5,91.6)	(89.3,90.8)	94.4	5.3	95.2	100.1
		200	100	6	(59.0,91.6)	(84.3,91.6)	95.0	19.5	96.4	99.1
				10	(82.7,91.6)	(87.0,91.6)	96.0	15.4	96.1	100.2
				20	(86.7,91.6)	(88.3,91.6)	93.7	11.1	95.0	100.0
			200	6	(59,91.6)	(85.8,91.6)	95.8	17.0	96.3	100.0
				10	(81.1,91.6)	(87.6,91.6)	94.5	12.9	95.2	100.2
				20	(86.7,91.6)	(88.8,91.2)	95.1	9.2	95.4	100.0
			500	6	(59.0,91.6)	(87.0,91.6)	95.1	13.4	95.6	100.4
				10	(81.8,91.6)	(88.3,91.5)	93.8	9.6	94.3	100.3
				20	(86.6,91.6)	(89.1,90.9)	94.4	6.7	94.7	100.1
			1000	6	(59.0,91.6)	(87.6,91.6)	95.4	11.0	95.8	100.2
				10	(83.3,91.6)	(88.9,91.2)	93.8	7.6	94.7	100.3
				20	(87.1,91.6)	(89.3,90.7)	95.0	5.0	95.0	100.1
$(1.2, 2)$	$(2, 2.6)$	50	100	6	(56.1,97.8)	(82.1,93.4)	93.9	25.9	96.0	98.2
				10	(75.5,95.0)	(85.5,92.2)	95.2	20.2	97.0	99.3
				20	(83.3,93.2)	(87.2,91.7)	94.6	14.3	95.5	100.1
			200	6	(56.6,97.4)	(83.4,92.8)	94.8	22.8	95.9	98.5
				10	(74.9,95.2)	(86.2,91.8)	95.6	16.4	96.8	99.9
				20	(83.4,93.4)	(87.8,91.3)	95.4	11.5	96.3	100.3
			500	6	(56.3,97.3)	(85.1,92.0)	95.1	18.1	97.0	99.5
				10	(76.9,95.2)	(87.3,91.4)	95.4	12.2	97.0	100.1
				20	(83.7,93.3)	(88.5,91.2)	94.4	8.1	95.3	100.5
			1000	6	(55.6,97.4)	(85.5,92.0)	95.4	15.3	95.7	99.7
				10	(75.0,95.2)	(87.7,91.2)	95.1	9.5	95.9	100.2
				20	(83.3,93.5)	(89.0,91.0)	95.4	6.1	95.2	100.4
		100	100	6	(58.0,96.2)	(83.8,92.9)	95.2	23.6	97.2	98.9
				10	(80.7,93.8)	(86.2,92.0)	94.4	18.0	96.9	99.2
				20	(85.6,92.4)	(87.7,91.5)	95.6	12.4	96.0	100.1
			200	6	(58.7,95.6)	(84.1,92.3)	95.0	20.1	96.8	98.7
				10	(82.0,94.0)	(86.7,91.7)	94.6	14.9	95.8	99.8
				20	(85.5,92.4)	(88.2,91.3)	94.3	10.2	95.8	100.3
			500	6	(59.0,96.6)	(86.2,91.8)	95.7	15.8	96.4	99.7
				10	(81.4,94.0)	(87.7,91.3)	96.2	11.3	96.6	100.1
				20	(85.3,92.5)	(88.7,91.1)	95.1	7.5	95.8	100.4
			1000	6	(58.5,95.4)	(86.1,91.4)	95.1	13.1	96.3	99.6
				10	(80.4,93.8)	(88.1,91.2)	94.7	9.0	95.4	100.2
				20	(85.6,92.4)	(89.1,91.0)	95.2	5.7	94.6	100.3
		200	100	6	(66.0,94.6)	(85.0,92.5)	95.0	20.2	96.4	98.7
				10	(84.2,92.8)	(86.9,91.7)	94.7	15.2	96.6	99.6
				20	(87.2,91.9)	(88.1,91.4)	94.5	10.5	95.8	100.1
			200	6	(59.2,94.4)	(85.8,92.0)	96.0	17.8	96.7	99.3
				10	(83.8,92.6)	(87.4,91.6)	94.5	13.1	95.8	99.9
				20	(87.0,91.7)	(88.5,91.2)	93.8	9.0	95.0	100.1
			500	6	(68.1,94.4)	(86.5,91.5)	95.2	14.0	96.8	99.7
		200		10	(83.9,92.7)	(88.1,91.3)	94.6	10.2	95.5	100.2
				20	(86.8,91.8)	(88.9,90.9)	93.9	6.8	94.4	100.1
			1000	6	(66.6,94.1)	(87.5,91.4)	95.0	11.5	95.8	100.1
				10	(84.3,92.7)	(88.6,91.1)	96.0	8.0	95.8	100.5
				20	(87.1,91.8)	(89.3,90.8)	95.0	5.3	95.4	100.2

Table A5: Simulation results: mean estimated success probability achieved under the estimated optimal package ($p_{\hat{\boldsymbol{x}}^{opt}}(\hat{\boldsymbol{\beta}})$), mean true success probability achieved under the estimated optimal package and ($p_{\hat{\boldsymbol{x}}^{opt}}(\boldsymbol{\beta}^{\star})$), and mean estimated success probability achieved under the true optimal package ($p_{\boldsymbol{x}^{opt}}(\hat{\boldsymbol{\beta}})$). Unit costs were $c_1 = 1$ and $c_2 = 8$. All results are multiplied by 100.

$\exp(\boldsymbol{\beta}^{\star})$	n_{1j}	n_{2j}	J	$\overline{p_{\hat{\boldsymbol{x}}^{opt}}(\hat{\boldsymbol{\beta}})}$ ($\times 100$)	$\overline{p_{\hat{\boldsymbol{x}}^{opt}}(\boldsymbol{\beta}^{\star})}$ ($\times 100$)	$\overline{p_{\boldsymbol{x}^{opt}}(\hat{\boldsymbol{\beta}})}$ ($\times 100$)
$(1, 2)$	50	100	6	89.98	88.86	89.61
			10	90.00	89.29	89.82
			20	90.00	89.63	89.86
		200	6	89.98	88.89	89.62
			10	90.00	89.51	89.79
			20	90.00	89.75	89.90
		500	6	89.96	89.42	89.70
			10	90.00	89.62	89.96
			20	90.00	89.83	89.94
		1000	6	89.99	89.48	89.81
			10	90.00	89.75	89.95
			20	90.00	89.91	89.96
	100	100	6	89.95	88.91	89.64
			10	90.00	89.51	89.86
			20	90.00	89.78	89.91
		200	6	90.00	89.25	89.90
			10	90.00	89.64	89.82
			20	90.00	89.83	89.95
		500	6	89.97	89.54	89.74
			10	90.00	89.75	89.86
			20	90.00	89.90	89.99
		1000	6	90.00	89.55	89.92
			10	90.00	89.83	89.91
			20	90.00	89.99	89.95
	200	100	6	90.00	89.28	89.84
			10	90.00	89.69	89.82
			20	90.00	89.87	89.92
		200	6	90.00	89.44	89.81
			10	90.00	89.71	89.88
			20	90.00	89.97	89.90
		500	6	90.00	89.59	89.87
			10	90.00	89.86	89.87
			20	90.00	89.94	89.98
		1000	6	90.00	89.71	89.79
			10	90.00	89.89	89.94
			20	90.00	89.99	89.97
$(1.2, 1.5)$	50	100	6	89.31	88.70	89.03
			10	89.72	89.45	89.60
			20	89.95	89.90	89.85
		200	6	89.34	89.14	88.93
			10	89.83	89.73	89.50
			20	89.98	90.01	89.87
		500	6	89.64	89.49	89.31
			10	89.92	89.85	89.86
			20	90.00	89.98	89.95
		1000	6	89.74	89.66	89.47
			10	89.94	89.97	89.76
			20	90.00	90.00	89.97
	100	100	6	89.49	88.99	89.34
			10	89.88	89.76	89.80
			20	89.98	89.94	89.93
		200	6	89.65	89.47	89.43
			10	89.92	89.88	89.79
			20	90.00	90.02	89.91
		500	6	89.87	89.64	89.78
			10	89.96	89.97	89.84
			20	90.00	90.06	89.91
		1000	6	89.88	89.77	89.65
			10	89.95	90.04	89.81
			20	90.00	90.01	89.97
	200	100	6	89.69	89.44	89.58
			10	89.93	89.88	89.88
			20	89.99	89.92	90.03
		200	6	89.85	89.78	89.69
			10	89.94	89.91	89.86
			20	90.00	89.97	89.99
		500	6	89.91	89.79	89.76
			10	89.99	89.99	89.89
			20	90.00	90.00	89.98
		1000	6	89.89	89.88	89.78
			10	89.99	90.03	89.91
			20	90.00	90.01	89.98

$\exp(\boldsymbol{\beta}^{\star})$	n_{1j}	n_{2j}	J	$\overline{p_{\hat{\boldsymbol{x}}^{opt}}(\hat{\boldsymbol{\beta}})}$ (×100)	$\overline{p_{\hat{\boldsymbol{x}}^{opt}}(\boldsymbol{\beta}^{\star})}$ (×100)	$\overline{p_{\boldsymbol{x}^{opt}}(\hat{\boldsymbol{\beta}})}$ (×100)
(1.2, 2)	50	100	6	89.92	88.72	89.04
			10	90.00	89.27	89.69
			20	90.00	89.68	89.81
		200	6	89.97	88.95	89.38
			10	90.00	89.46	89.65
			20	90.00	89.84	89.83
		500	6	90.00	89.32	89.37
			10	90.00	89.72	89.73
			20	90.00	89.94	89.83
		1000	6	90.00	89.41	89.62
			10	90.00	89.82	89.78
			20	90.00	90.00	89.89
	100	100	6	90.00	89.08	89.57
			10	90.00	89.47	89.78
			20	90.00	89.77	89.90
		200	6	89.97	89.13	89.53
			10	90.00	89.63	89.76
			20	90.00	89.87	89.89
		500	6	90.00	89.53	89.63
			10	90.00	89.76	89.82
			20	90.00	90.00	89.87
		1000	6	90.00	89.59	89.67
			10	90.00	89.90	89.87
			20	90.00	90.00	89.92
	200	100	6	90.00	89.26	89.54
			10	90.00	89.61	89.86
			20	90.00	89.92	89.93
		200	6	90.00	89.44	89.76
			10	90.00	89.75	89.90
			20	90.00	89.92	89.97
		500	6	90.00	89.58	89.76
			10	90.00	89.86	89.88
			20	90.00	89.97	89.97
		1000	6	90.00	89.72	89.73
			10	90.00	89.97	89.86
			20	90.00	90.00	89.98

Table A6: Simulation results for estimated optimal package. Analogue of Table 2 in the main paper and Table A3 here when unit costs were $c_1 = 1$ and $c_2 = 6$. Results include bias for optimal individual components (times 100) and empirical mean squared error (times 100) of the second-stage recommended intervention and the final estimated optimal package.

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	Stage 1 Bias$_1$ (×100)	Stage 1 Bias$_2$ (×100)	Stage 1 RMSE (×100)	Stage 2 Bias$_1$ (×100)	Stage 2 Bias$_2$ (×100)	Stage 2 RMSE (×100)
Scenario 1 ($J_1 = J_2 = J$)										
$(1, 2)$	$(0, 3.2)$	50	100	6	68.5	-46.0	166.6	47.3	-12.8	107.3
				10	56.1	-19.2	126.6	33.5	-7.3	86.3
				20	41.8	-8.4	100.3	22.6	-3.3	69.8
			500	6	69.6	-47.0	166.8	32.7	-6.7	86.9
				10	61.4	-18.6	129.8	19.2	-3.5	64.3
				20	41.6	-9.9	100.2	8.2	-0.9	41.9
		100	100	6	63.5	-29.7	146.1	42.6	-9.3	99.7
				10	44.6	-13.2	106.9	27.4	-5.4	77.8
				20	24.6	-4.3	76.7	13.5	-1.5	54.3
			500	6	64.4	-31.5	147.2	30.4	-5.7	82.2
				10	51.9	-10.8	114.6	15.9	-2.0	58.4
				20	25.9	-5.6	78.1	3.6	-0.2	28.4
$(1.2, 1.5)$	$(2, 4.5)$	50	100	6	-71.8	-87.7	210.0	-40.2	-6.5	108.3
				10	-53.2	-35.7	145.8	-30.1	2.8	85.7
				20	-35.5	-9.0	99.8	-12.5	3.5	56.9
			500	6	-73.2	-90.3	211.4	-27.2	0.6	85.9
				10	-54.7	-35.6	146.9	-12.5	3.2	56.3
				20	-36.8	-8.9	100.6	-4.3	2.5	32.9
		100	100	6	-58.6	-56.5	176.2	-36.1	-4.1	101.2
				10	-37.6	-14.4	108.8	-17.6	3.1	67.6
				20	-21.0	-2.2	76.7	-6.3	1.8	42.0
			500	6	-56.1	-57.3	173.6	-23.4	3.2	75.5
				10	-38.8	-17.0	113.7	-8.9	3.8	47.2
				20	-22.2	-1.7	78.0	-1.4	2.1	21.6
$(1.2, 2)$	$(2, 2.6)$	50	100	6	-90.3	-26.6	170.4	-80.5	5.3	135.0
				10	-79.7	-5.1	141.4	-68.2	7.2	120.6
				20	-64.9	1.7	119.6	-51.6	7.6	104.4
			500	6	-87.8	-26.0	168.2	-64.9	6.8	119.1
				10	-78.7	-5.5	140.0	-48.9	7.0	101.7
				20	-66.0	2.7	120.3	-31.2	6.0	81.3
			100	6	-81.1	-15.4	152.8	-67.9	5.9	122.5
				10	-64.3	-0.7	122.3	-54.1	5.1	107.2
				20	-53.0	4.2	106.3	-40.5	6.3	92.4
			500	6	-81.8	-11.1	151.5	-58.0	6.8	111.2
				10	-65.3	0.1	123.1	-40.7	6.1	92.6
				20	-52.2	4.2	105.7	-24.2	4.9	71.5
Scenario 2a ($J_1 = 6,\ J_2 = 12$)										
$(1, 2)$	$(0, 3.2)$	50	200		70.7	-43.8	168.2	29.5	-6.2	81.6
$(1.2, 1.5)$	$(2, 4.5)$				-69.4	-95.4	214.4	-24.9	0.5	79.9
$(1.2, 2)$	$(2, 2.6)$				-91.5	-32.9	173.5	-62.0	5.9	115.6
Scenario 2b ($J_1 = 10,\ J_2 = 20$)										
$(1, 2)$	$(0, 3.2)$	50	200		64.5	-18.6	133.3	33.5	-5.0	83.8
$(1.2, 1.5)$	$(2, 4.5)$				-47.8	-35.1	141.8	-10.0	2.4	52.5
$(1.2, 2)$	$(2, 2.6)$				-69.0	-6.6	134.6	-33.4	3.9	84.2

Table A7: Simulation results for estimated optimal package. Analogue of Table 3 in the main paper and Table A4 here when unit costs were $c_1 = 1$ and $c_2 = 6$. Results include empirical 2.5% and 97.5% quantiles of true success rate of second-stage recommended intervention (PrOpt1) and final estimated optimal intervention (PrOpt2), coverage rate of 95% confidence set for optimal intervention package (SetCP95), and its relative size (SetPerc%). Results also include the coverage rate of 95% confidence bands for outcome probability under all values on a grid of $\mathcal{X}$ (BandsCP95) and the ratio between the mean cost of the estimated optimal intervention and the cost of the optimal true intervention (RelCost).

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	PrOpt1 (Q2.5,Q97.5)	PrOpt2 (Q2.5,Q97.5)	SetCP95	SetPerc%	BandsCP95	RelCost
Scenario 1 ($J_1 = J_2 = 20$)										
$(1, 2)$	$(0, 3.2)$	50	100	6	(50.0,97.0)	(82.5,93.6)	92.9	16.4	97.9	98.4
				10	(77.1,96.0)	(85.8,92.4)	94.8	11.3	97.4	99.5
				20	(83.6,93.9)	(87.3,91.9)	94.0	7.5	97.0	100.2
			500	6	(50.0,97.0)	(85.0,92.4)	94.0	10.2	97.4	99.6
				10	(78.1,95.9)	(87.0,91.6)	94.6	6.4	96.9	99.9
				20	(83.5,93.8)	(88.3,91.2)	94.8	4.0	97.4	100.1
		100	100	6	(50.0,97.0)	(84.5,93.1)	93.7	14.0	97.0	99.3
				10	(81.7,94.3)	(86.4,92.3)	94.6	9.4	97.9	99.7
				20	(85.2,93.2)	(87.8,91.7)	94.9	6.2	96.8	100.2
			500	6	(50.0,97.0)	(86.0,92.1)	94.2	9.0	97.0	99.8
				10	(81.7,94.9)	(87.7,91.5)	95.4	5.8	96.8	100.2
				20	(85.6,92.9)	(88.9,91.1)	95.3	3.7	97.7	100.1
$(1.2, 1.5)$	$(2, 4.5)$	50	100	6	(56.9,91.6)	(81.3,91.6)	94.0	24.4	95.8	97.3
				10	(68.9,91.6)	(85.2,91.6)	94.4	18.4	96.2	99.5
				20	(81.0,91.6)	(87.3,91.6)	95.0	13.2	96.2	100.3
			500	6	(56.7,91.6)	(84.7,91.6)	94.8	17.4	96.8	99.2
				10	(68.4,91.6)	(87.0,91.6)	95.5	11.8	95.6	100.2
				20	(81.8,91.6)	(88.7,91.3)	95.1	7.6	95.9	100.4
		100	100	6	(58.6,91.6)	(83.0,91.6)	94.0	21.9	96.8	97.9
				10	(79.4,91.6)	(86.4,91.6)	94.8	16.7	96.3	100.0
				20	(84.6,91.6)	(87.8,91.6)	94.9	12.2	95.5	100.2
			500	6	(59.0,91.6)	(85.6,91.6)	94.3	15.4	95.5	99.9
				10	(78.5,91.6)	(87.7,91.6)	95.5	10.6	95.8	100.5
				20	(83.9,91.6)	(89.0,91.0)	95.4	7.1	95.7	100.4
$(1.2, 2)$	$(2, 2.6)$	50	100	6	(56.1,97.8)	(82.1,93.3)	94.2	25.9	96.1	97.3
				10	(75.5,95.0)	(85.6,92.2)	95.4	20.1	97.0	98.6
				20	(83.2,93.1)	(87.1,91.6)	95.0	14.2	95.6	99.7
			500	6	(56.3,97.3)	(85.0,91.9)	95.1	18.1	96.9	98.7
				10	(76.9,95.2)	(87.3,91.2)	95.4	12.2	96.8	99.6
				20	(83.7,93.3)	(88.3,91.0)	94.6	8.1	95.4	100.3
		100	100	6	(58.0,96.2)	(83.8,92.9)	95.4	23.4	97.2	98.2
				10	(80.7,93.8)	(86.2,91.9)	94.6	17.9	96.8	98.7
				20	(85.6,92.4)	(87.6,91.4)	95.7	12.4	95.8	99.8
			500	6	(59.0,96.6)	(86.2,91.7)	95.5	15.8	96.4	99.0
				10	(81.4,94.0)	(87.6,91.2)	96.3	11.3	96.5	99.8
				20	(85.2,92.3)	(88.5,90.9)	95.4	7.4	95.8	100.3
Scenario 2a ($J_1 = 6,\ J_2 = 12$)										
$(1, 2)$	$(0, 3.2)$	50	200		(50,97)	(85.7,92.5)	94.2	9.7	97.0	99.6
$(1.2, 1.5)$	$(2, 4.5)$				(56.3,91.6)	(85.4,91.6)	94.6	17.5	96.4	99.3
$(1.2, 2)$	$(2, 2.6)$				(55.7,97.9)	(85.2,91.9)	95.0	17.4	95.8	98.5
Scenario 2b ($J_1 = 10,\ J_2 = 20$)										
$(1, 2)$	$(0, 3.2)$	50	200		(77.5,95.8)	(87.1,91.5)	94.6	6.4	97.2	99.7
$(1.2, 1.5)$	$(2, 4.5)$				(71.3,91.6)	(87.2,91.6)	95.2	12.0	95.7	100.2
$(1.2, 2)$	$(2, 2.6)$				(75.9,95.6)	(87.1,91.4)	95.8	12.6	96.8	99.9

Table A8: Simulation results for estimated optimal package under nonlinear cost function $C(\boldsymbol{x}) = x_1 + 5x_2 + 3\exp(-0.5x_2)$. Results include bias for optimal individual components (times 100) and empirical mean squared error (times 100) of the second-stage recommended package and the final estimated optimal package.

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	Stage 1 Bias$_1$ (×100)	Stage 1 Bias$_2$ (×100)	Stage 1 RMSE (×100)	Stage 2 Bias$_1$ (×100)	Stage 2 Bias$_2$ (×100)	Stage 2 RMSE (×100)
$(1,2)$	$(0,3.2)$	50	100	6	69.3	-47.7	164.8	40.5	-9.5	94.8
				10	52.4	-15.5	120.4	29.9	-5.1	75.6
				20	31.3	-6.2	84.0	16.6	-1.6	52.5
			200	6	65.7	-43.7	161.5	32.5	-7.7	83.6
				10	54.3	-16.2	121.3	22.4	-3.5	63.6
				20	29.8	-5.4	82.2	8.3	-0.4	34.1
			500	6	71.7	-41.8	167.3	26.3	-5.6	72.5
				10	52.2	-14.0	120.2	14.4	-1.1	48.4
				20	33.3	-6.1	86.1	5.8	1.1	25.3
			1000	6	67.2	-37.2	161.0	18.2	-2.9	58.3
				10	52.2	-16.5	120.2	10.2	-0.4	39.0
				20	32.6	-7.4	85.4	3.2	0.9	16.5
		100	100	6	57.2	-26.5	137.8	35.1	-8.1	86.0
				10	39.2	-8.5	98.4	23.2	-4.1	66.1
				20	15.0	-1.5	56.0	8.2	-0.1	35.1
			200	6	56.7	-29.0	138.8	28.5	-5.4	75.7
				10	38.5	-10.1	96.6	16.0	-1.6	52.6
			200	20	18.7	-1.8	61.2	5.4	0.9	25.4
			500	6	57.1	-28.8	141.4	21.0	-3.1	61.6
				10	37.4	-10.7	96.3	12.1	-0.9	44.2
				20	18.2	-1.6	60.2	3.6	1.3	17.0
			1000	6	56.4	-28.3	139.1	16.0	-2.0	53.0
				10	40.7	-7.3	98.3	7.5	0.4	31.3
				20	18.2	-3.6	61.5	3.2	0.8	15.8
		200	100	6	45.4	-17.8	116.3	27.4	-5.9	75.7
				10	26.6	-3.9	76.6	14.1	-1.4	50.0
				20	9.6	-0.1	41.2	5.0	0.5	25.2
			200	6	43.9	-18.5	116.2	21.7	-2.5	64.2
				10	25.3	-3.6	73.8	11.4	-0.3	43.1
				20	8.7	0.5	39.4	4.2	1.0	21.0
			500	6	44.7	-15.6	114.9	15.6	-0.5	51.9
				10	25.4	-4.5	74.8	7.5	0.4	31.8
				20	6.9	0.5	35.4	2.9	1.2	15.4
			1000	6	45.6	-16.6	115.7	12.3	-0.6	44.7
				10	27.7	-3.8	77.7	5.1	0.9	23.7
				20	7.8	-0.3	37.6	2.2	1.1	10.8

Table A.8: continued

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	Stage 1 Bias$_1$ (×100)	Stage 1 Bias$_2$ (×100)	Stage 1 RMSE (×100)	Stage 2 Bias$_1$ (×100)	Stage 2 Bias$_2$ (×100)	Stage 2 RMSE (×100)
(1.2, 1.5)	(1.9, 4.6)	50	100	6	-61.9	-98.7	214.2	-40.0	-8.0	108.8
				10	-52.2	-41.0	145.3	-26.8	-1.1	83.8
				20	-32.3	-11.0	96.8	-14.0	1.0	61.1
			200	6	-63.9	-94.1	209.3	-31.3	-3.1	93.5
				10	-48.6	-37.6	140.7	-20.0	0.8	73.2
				20	-33.5	-11.9	98.1	-8.2	0.3	49.2
			500	6	-69.3	-98.2	214.8	-24.5	-1.5	83.9
				10	-49.4	-39.4	140.7	-12.7	0.3	58.8
				20	-33.9	-12.2	99.5	-4.2	-0.0	38.3
			1000	6	-63.2	-100.6	214.9	-17.0	-0.3	70.3
				10	-47.2	-40.0	142.4	-6.3	1.1	45.1
				20	-32.1	-14.5	98.1	0.7	-1.7	23.7
		100	100	6	-55.4	-67.7	182.0	-32.1	-5.0	95.4
				10	-40.6	-21.5	114.1	-18.6	-0.5	71.1
				20	-22.1	-5.7	78.7	-5.8	0.0	45.8
			200	6	-55.8	-64.2	175.6	-24.6	-2.5	82.4
				10	-37.1	-18.4	110.8	-16.5	1.2	65.4
				20	-22.9	-5.3	79.9	-4.3	-1.1	40.4
			500	6	-57.8	-64.8	178.5	-19.3	0.9	72.3
				10	-38.3	-18.3	114.9	-7.7	1.0	48.6
				20	-23.0	-5.7	80.0	-0.4	-1.3	29.1
			1000	6	-53.1	-65.1	177.9	-11.4	-1.1	62.0
				10	-35.9	-20.0	111.1	-3.2	-0.3	38.4
				20	-20.7	-6.7	77.4	1.7	-1.6	19.2
		200	100	6	-44.4	-34.4	139.8	-21.8	-3.2	80.7
				10	-25.7	-8.7	89.2	-13.3	-0.4	61.7
				20	-9.8	-1.4	57.0	-2.2	-0.1	36.6
			200	6	-45.7	-35.9	142.6	-19.4	-0.2	73.2
				10	-24.5	-8.4	87.1	-8.3	-0.8	51.2
				20	-11.0	-2.2	58.8	-1.5	-1.9	33.4
			500	6	-44.3	-33.7	140.6	-15.7	1.6	63.7
				10	-27.8	-8.5	89.4	-5.7	0.0	43.2
				20	-9.2	-3.7	56.3	0.4	-1.9	26.5
			1000	6	-47.1	-39.6	148.8	-11.7	0.3	55.4
				10	-26.2	-10.6	90.2	-2.8	-0.6	35.2
				20	-10.4	-2.8	58.5	2.5	-2.4	15.5
(1.2, 2)	(1.8, 2.7)	50	100	6	-77.8	-35.5	164.7	-74.0	4.5	125.7
				10	-71.6	-4.6	133.8	-67.7	8.0	116.6
				20	-65.0	4.5	116.8	-59.7	10.1	107.9
			200	6	-80.3	-27.6	161.0	-74.0	6.5	124.2
				10	-70.5	-5.2	132.8	-67.7	10.0	115.0
				20	-64.0	3.2	116.0	-55.7	9.8	103.3
			500	6	-76.1	-29.4	162.0	-65.4	7.3	115.9
				10	-68.6	-6.2	131.2	-59.3	9.9	106.9
				20	-63.2	2.8	116.0	-50.2	10.1	97.1
			1000	6	-78.6	-27.1	162.6	-62.7	8.9	111.6
				10	-70.8	-6.8	134.2	-56.2	10.5	104.1
				20	-62.7	3.3	115.0	-42.3	9.2	89.0
		100	100	6	-74.1	-16.9	146.2	-69.1	7.1	119.5
				10	-65.6	2.6	121.2	-62.3	9.0	111.3
				20	-54.6	5.3	104.6	-48.6	8.5	97.8
			200	6	-72.0	-16.3	146.4	-62.2	6.5	113.5
				10	-69.1	1.5	122.7	-60.4	9.9	108.7
				20	-55.9	4.6	105.5	-45.5	8.2	93.5
			500	6	-73.2	-17.5	147.6	-65.4	8.9	114.5
				10	-63.1	1.6	120.4	-53.5	9.4	101.6
				20	-57.7	5.9	107.1	-45.5	9.5	92.0
			1000	6	-70.7	-14.5	141.8	-58.1	8.8	107.3
				10	-67.9	2.3	122.6	-50.9	10.2	98.7
				20	-58.1	6.4	107.0	-34.6	7.5	81.6
		200	100	6	-68.3	-9.2	132.7	-63.4	5.4	113.2
				10	-58.6	5.0	110.1	-52.0	8.4	101.3
				20	-49.2	6.8	97.0	-42.4	7.7	89.9
			200	6	-66.4	-7.7	130.7	-59.0	7.0	109.0
				10	-59.4	5.7	109.7	-51.2	8.7	99.8
				20	-49.4	7.0	97.9	-42.1	8.7	89.1
			500	6	-67.5	-7.5	131.3	-55.1	8.7	104.6
				10	-57.4	4.5	109.5	-50.2	9.8	97.5
				20	-48.6	7.3	97.0	-37.8	8.3	84.6
			1000	6	-68.2	-8.4	132.3	-54.4	9.4	102.2
				10	-63.2	6.1	113.6	-48.4	9.6	95.3
				20	-47.9	6.1	96.3	-29.2	6.8	74.8

Table A9: Simulation results for estimated optimal package , under nonlinear cost function $C(\boldsymbol{x}) = x_1 + 5x_2 + 3\exp(-0.5x_2)$. Analogue of Table 3 in the main paper and Table A4. Results include empirical 2.5% and 97.5% quantiles of true success rate of second-stage recommended intervention (PrOpt1) and final estimated optimal intervention (PrOpt2), coverage rate of 95% confidence set for optimal intervention package (SetCP95), and its relative size (SetPerc%). Results also include the coverage rate of 95% confidence bands for outcome probability under all values on a grid of $\mathcal{X}$ (BandsCP95) and the ratio between the mean cost of the estimated optimal intervention and the cost of the optimal true intervention (RelCost).

$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	PrOpt1 (Q2.5,Q97.5)	PrOpt2 (Q2.5,Q97.5)	SetCP95	SetPerc%	BandsCP95	RelCost
$(1,2)$	$(0,3.2)$	50	100	6	(50.0,97.0)	(83.1,93.7)	93.0	29.5	96.0	99.8
				10	(78.9,96)	(85.8,92.4)	93.2	20.6	95.8	100.4
				20	(84.1,94.1)	(87.4,91.9)	93.6	13.5	95.6	100.6
			200	6	(50.0,97.0)	(84.1,92.9)	92.8	24.3	95.4	99.9
				10	(78.8,96.0)	(86.7,92.4)	94.0	16.6	95.2	100.4
				20	(84.1,94.5)	(88.2,91.9)	93.2	10.8	95.5	100.4
			500	6	(50.0,97.0)	(85.0,92.4)	92.7	19.1	95.5	100.0
				10	(77.7,96.0)	(87.4,91.9)	93.3	12.6	95.8	100.6
				20	(84.1,93.7)	(88.9,91.3)	92.0	8.0	94.8	100.7
			1000	6	(50.0,97.0)	(86.7,91.9)	93.3	15.5	96.2	100.3
				10	(77.7,96.5)	(88.2,91.3)	93.9	9.9	96.2	100.5
				20	(84.1,94.1)	(89.6,90.8)	91.4	6.2	95.9	100.5
		100	100	6	(53.4,97.0)	(84.1,93.3)	92.8	24.6	95.0	99.9
				10	(82.1,95.2)	(86.7,92.4)	92.7	17.4	94.4	100.3
				20	(85.8,93.3)	(88.2,91.9)	93.9	11.4	94.6	100.5
			200	6	(50.0,96.3)	(85.8,92.9)	93.8	20.9	96.6	100.2
				10	(82.1,94.5)	(87.4,91.9)	94.2	14.2	95.4	100.5
				20	(85.8,93.3)	(88.9,91.9)	94.2	9.4	94.6	100.6
			500	6	(50.0,96.8)	(86.7,91.9)	94.4	16.2	95.6	100.4
				10	(81.1,94.5)	(88.2,91.9)	93.6	10.8	95.4	100.5
				20	(86.6,93.3)	(89.6,91.3)	94.2	7.1	95.7	100.6
			1000	6	(50.0,96.8)	(86.7,91.9)	93.6	13.0	95.6	100.4
				10	(83.1,94.8)	(88.9,91.3)	92.3	8.7	95.2	100.6
				20	(85.8,93.3)	(89.6,90.8)	90.4	5.6	96.5	100.4
		200	100	6	(69.7,96.0)	(85.8,92.9)	93.7	21.0	95.1	100.0
				10	(85.0,93.7)	(87.4,92.4)	92.9	14.5	94.9	100.5
				20	(87.4,92.4)	(88.9,91.9)	94.2	9.6	96.0	100.4
			200	6	(68.2,95.5)	(86.7,92.4)	93.4	17.5	95.2	100.6
				10	(85.0,93.3)	(88.2,91.9)	94.0	12.1	95.4	100.6
				20	(87.4,92.9)	(88.9,91.3)	92.6	8.1	94.9	100.6
			500	6	(73.9,95.8)	(87.4,91.9)	95.1	13.6	95.4	100.8
				10	(85.0,93.7)	(88.9,91.3)	93.2	9.3	95.1	100.6
				20	(87.4,92.4)	(89.6,91.3)	91.8	6.2	95.4	100.5
			1000	6	(72.5,95.5)	(88.2,91.9)	92.9	11.1	95.4	100.6
				10	(85.0,93.3)	(88.9,91.3)	93.1	7.6	94.5	100.6
				20	(87.4,92.4)	(89.6,90.8)	91.1	5.0	95.6	100.4

Table A.9: continued

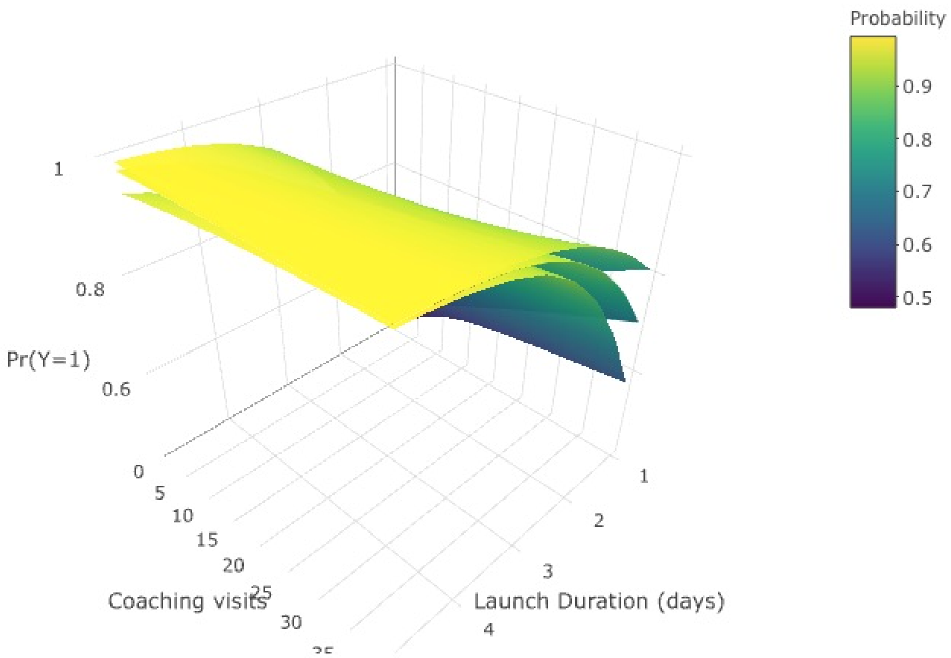
$e^{\boldsymbol{\beta}^\star}$	$\boldsymbol{x}^{opt}$	n_{1j}	n_{2j}	J	PrOpt1 (Q2.5,Q97.5)	PrOpt2 (Q2.5,Q97.5)	SetCP95	SetPerc%	BandsCP95	RelCost
$(1.2, 1.5)$	$(1.9, 4.6)$	50	100	6	(56.3,91.6)	(82.4,91.6)	94.3	27.4	95.2	97.0
				10	(71.8,91.6)	(85.6,91.6)	93.6	19.2	95.1	98.8
				20	(81.8,91.6)	(87.1,91.6)	94.4	13.1	94.8	99.7
			200	6	(57.2,91.6)	(84.1,91.6)	95.2	23.0	95.0	98.2
				10	(72.6,91.6)	(87.0,91.6)	95.6	15.6	96.4	99.4
				20	(81.8,91.6)	(87.9,91.6)	95.0	10.4	95.4	99.7
			500	6	(56.8,91.6)	(85.1,91.6)	95.1	18.7	94.7	98.8
				10	(71.8,91.6)	(87.1,91.6)	95.2	11.7	96.0	99.6
				20	(82.4,91.6)	(88.5,91.3)	94.5	7.4	94.9	99.8
			1000	6	(56.8,91.6)	(86.4,91.6)	95.0	15.1	95.0	99.3
				10	(71.7,91.6)	(88.1,91.6)	94.8	9.1	95.7	100.0
				20	(81.8,91.6)	(89.2,91.0)	93.6	5.5	94.8	99.7
		100	100	6	(59.0,91.6)	(84.1,91.6)	95.0	23.8	95.9	97.8
				10	(79.2,91.6)	(86.6,91.6)	94.0	17.1	95.0	99.2
				20	(84.1,91.6)	(87.9,91.6)	94.2	12.0	95.4	99.8
			200	6	(59.0,91.6)	(85.1,91.6)	94.4	20.0	95.6	98.6
				10	(79.6,91.6)	(87.1,91.6)	95.1	13.9	95.6	99.6
				20	(84.1,91.6)	(88.3,91.5)	94.4	9.7	95.9	99.6
			500	6	(59.0,91.6)	(86.1,91.6)	94.8	15.7	94.8	99.4
				10	(79.2,91.6)	(87.9,91.6)	94.6	10.4	95.0	99.9
				20	(84.3,91.6)	(89.0,91.2)	94.0	6.9	95.4	99.7
			1000	6	(59.0,91.6)	(87.0,91.6)	95.0	13.4	95.1	99.4
				10	(79.2,91.6)	(88.4,91.5)	94.2	8.1	95.8	99.8
				20	(84.1,91.6)	(89.4,90.8)	93.8	5.1	95.7	99.8
		200	100	6	(64.8,91.6)	(85.1,91.6)	94.0	20.4	95.6	98.5
				10	(81.8,91.6)	(86.6,91.6)	94.4	15.2	95.3	99.4
				20	(86.1,91.6)	(88.3,91.6)	94.9	10.7	95.8	99.9
			200	6	(59.0,91.6)	(85.6,91.6)	95.0	17.3	95.2	99.2
				10	(82.4,91.6)	(87.7,91.6)	93.8	12.7	96.2	99.5
				20	(86.3,91.6)	(88.6,91.3)	93.8	9.0	94.8	99.6
			500	6	(59.0,91.6)	(87.1,91.6)	94.2	13.4	94.2	99.7
				10	(82.4,91.6)	(88.3,91.6)	94.4	9.4	94.8	99.8
				20	(86.3,91.6)	(89.2,91.0)	93.8	6.5	95.0	99.6
			1000	6	(59.0,91.6)	(87.3,91.6)	95.2	11.5	96.3	99.6
				10	(82.2,91.6)	(88.6,91.2)	94.6	7.5	95.6	99.8
				20	(86.1,91.6)	(89.4,90.8)	91.6	4.9	94.6	99.6
$(1.2, 2)$	$(1.8, 2.7)$	50	100	6	(55.9,97.0)	(82.4,93.3)	94.8	30.8	95.5	96.8
				10	(76.8,95.3)	(85.8,92.0)	95.5	21.3	95.6	98.1
				20	(84.1,93.4)	(87.4,91.5)	95.6	14.3	95.2	99.2
			200	6	(57.2,96.8)	(84.1,92.5)	94.8	25.5	95.8	97.4
				10	(78.0,95.3)	(86.7,91.7)	95.2	17.4	94.6	98.7
				20	(83.1,93.3)	(87.8,91.2)	94.2	11.3	95.4	99.4
			500	6	(56.8,97.0)	(85.1,92.0)	94.4	19.9	95.0	98.1
				10	(76.5,94.9)	(87.4,91.2)	96.0	12.9	95.5	99.2
				20	(83.4,93.3)	(88.6,90.8)	95.6	8.0	95.2	99.8
			1000	6	(56.3,97.9)	(85.8,91.3)	95.8	16.3	96.4	98.7
				10	(75.2,95.2)	(87.6,90.9)	94.2	10.1	94.8	99.6
				20	(83.6,93.4)	(88.9,90.7)	95.1	6.0	95.1	100.0
		100	100	6	(58.1,95.7)	(84.3,92.7)	94.9	25.8	95.6	97.8
				10	(82.1,94.2)	(86.7,91.9)	94.0	18.3	94.6	98.7
				20	(85.8,92.4)	(87.6,91.3)	95.0	12.3	95.2	99.4
			200	6	(58.1,96.4)	(85.8,92.3)	95.3	22.1	95.9	98.1
				10	(81.4,93.7)	(87.4,91.5)	95.9	15.2	95.8	99.1
				20	(85.8,92.4)	(88.2,91.1)	95.3	10.0	95.4	99.5
			500	6	(59.0,96.1)	(85.8,91.7)	94.8	17.1	95.4	98.5
				10	(81.4,93.8)	(87.8,91.2)	96.0	11.3	95.5	99.4
				20	(85.8,92.5)	(88.7,90.8)	95.4	7.3	95.2	99.9
			1000	6	(59.0,95.8)	(86.7,91.3)	95.3	13.8	94.9	98.9
				10	(82.1,93.9)	(88.2,90.9)	95.2	8.7	94.6	99.8
				20	(85.8,92.4)	(88.9,90.6)	94.2	5.6	94.2	100.0
		200	100	6	(65.1,94.6)	(85.2,92.3)	94.6	21.6	95.4	97.7
				10	(84.8,92.5)	(87.4,91.6)	96.1	15.3	95.7	99.2
				20	(87.4,91.6)	(88.2,91.2)	95.4	10.2	95.3	99.6
			200	6	(69.6,94.8)	(86.1,91.9)	95.2	18.4	94.8	98.4
				10	(85.0,92.9)	(87.6,91.5)	95.4	12.9	95.2	99.3
				20	(86.9,91.7)	(88.6,91.1)	95.3	8.7	95.4	99.9
			500	6	(63.5,94.4)	(86.9,91.4)	95.1	14.5	95.6	99.1
				10	(84.1,92.9)	(88.2,91.1)	95.3	9.9	94.8	99.7
				20	(87.4,91.7)	(88.9,90.8)	95.3	6.6	96.0	100.1
			1000	6	(59.0,94.2)	(87.6,91.2)	96.2	11.8	95.6	99.4
				10	(85.0,92.9)	(88.4,90.8)	95.4	7.8	95.8	99.8
				20	(86.7,91.6)	(89.1,90.6)	94.6	5.1	94.4	100.1

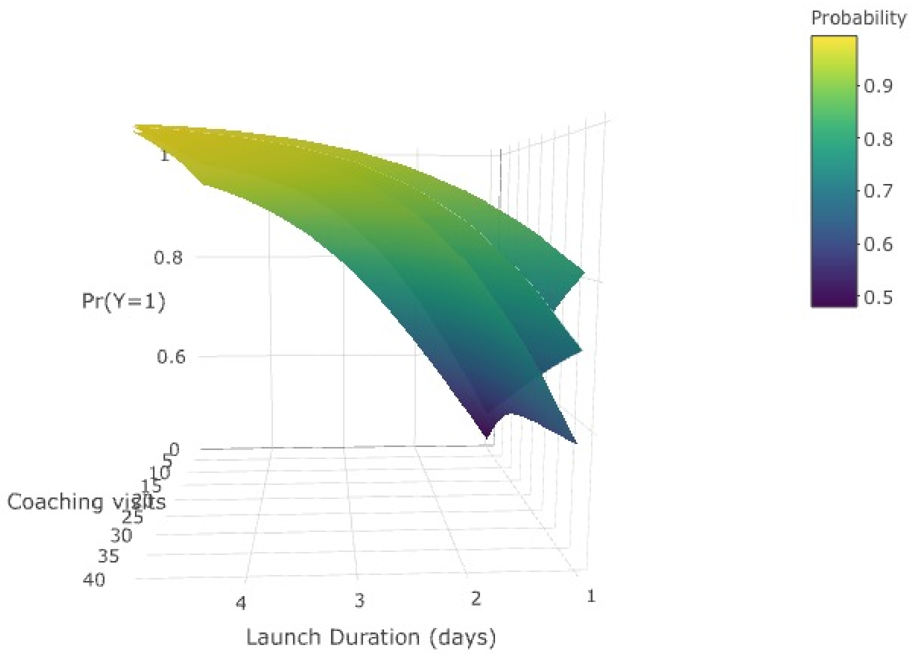
6 Confidence bands for oxytocin administration in the BetterBirth Study

Figure 2 below presents 95% confidence bands for the probability of oxytocin administration in the BetterBirth Study under all possible values of the intervention for a center with average monthly birth rate ($z = 175$). These confidence bands have an asymptotic simultaneous coverage of the probability of oxytocin administration under all intervention values. The figure includes four panels showing the bands from different angles. The middle curve is the estimated probability of oxytocin administration, $\hat{p}_{\boldsymbol{x}}(\hat{\boldsymbol{\beta}}; \boldsymbol{z} = 175)$.

References

Scheffé, H. (1959). *The analysis of variance*, Volume 72. John Wiley & Sons.


Probability
0.9
0.8
0.7
0.6
0.5
1
0.8
Pr(Y=1)
0.6
0
5
10
15
20
25
30
Coaching visits
1
2
3
4
Launch Duration (days)


Probability
0.9
0.8
0.7
0.6
0.5
1
0.8
Pr(Y=1)
0.6
0
5
10
15
20
25
30
35
40
Coaching visits
4
3
2
1
Launch Duration (days)

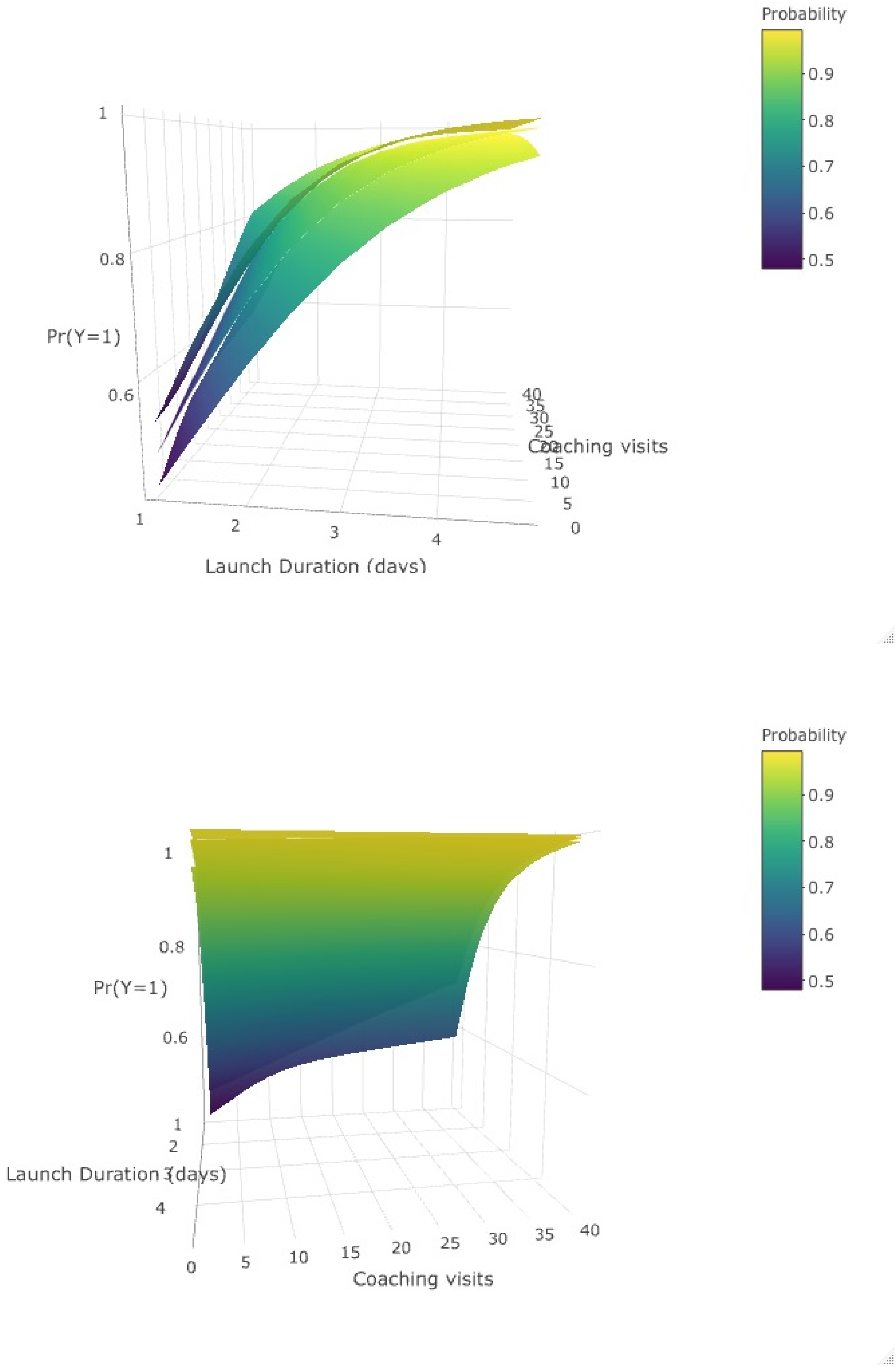


Figure 2: Confidence bands for probability of oxytocin use under various intervention compositions, for a center with average monthly number of births ($z = 175$). All four panels show the same confidence bands, looking from a different angle on the plot. The figure portrays three curves: the middle one is the estimated probability of success and the upper and lower curves are the boundaries of the confidence bands.