A model of a randomized experiment with an application to the PROWESS clinical trial

Amanda E. Kowalski

A Model of a Randomized Experiment with an Application to the PROWESS Clinical Trial[*]

Amanda E. Kowalski

March 13, 2019

Abstract

I develop a model of a randomized experiment with a binary intervention and a binary outcome. Potential outcomes in the intervention and control groups give rise to four types of participants. Fixing ideas such that the outcome is mortality, some participants would live regardless, others would be saved, others would be killed, and others would die regardless. These potential outcome types are not observable. However, I use the model to develop estimators of the number of participants of each type. The model relies on the randomization within the experiment and on deductive reasoning. I apply the model to an important clinical trial, the PROWESS trial, and I perform a Monte Carlo simulation calibrated to estimates from the trial. The reduced form from the trial shows a reduction in mortality, which provided a rationale for FDA approval. However, I find that the intervention killed two participants for every three it saved.

1 Introduction

Consider a randomized experiment with a binary intervention and a binary outcome. To fix ideas, suppose that the outcome is mortality. At the end of the experiment, we observe whether each participant is alive or dead. We also observe whether each participant was randomized into the intervention group or the control group. Using these data, we can calculate the reduced form, which is equal to the fraction of participants observed dead in the intervention group minus the fraction of participants observed dead in the control group. If the reduced form is negative, then we conclude that, on average, the intervention reduced mortality. If the intervention was access to a medical treatment, then the reduced form provides a rationale to expand access to the treatment.

However, there is a possibility that even though the intervention reduces mortality on average, some individuals would be killed by it. To be precise, within the experiment, the participants who would be killed are those that would die if they were assigned to the intervention group but would live if they were assigned to the control group. If we could identify such individuals, then their prevalence might provide a rationale against expanding access to the treatment.

However, individuals who would be killed are difficult to identify on the individual level because it is not possible to observe what mortality would have been in the control group for a participant assigned to the intervention group and vice versa. This problem is well-known, and it has been formalized with the concept of "potential outcomes" (Rubin, 1974, 1977; Holland, 1986). In the context of the experiment, each participant has a potential outcome in the control group and a potential outcome in the intervention group, but only one potential outcome is observed for each participant. On the basis of all combinations of potential outcomes, there are four types of participants in the experiment. In addition to those who would be killed, there are those who would be saved, those who would live regardless, and those who would die regardless.

The reduced form gives the difference between the fraction of participants who would be killed and the fraction of participants who would be saved. To see this, recall that the reduced form is equal to the fraction of participants observed dead in the intervention group minus the fraction of participants observed dead in the control group. In terms of potential outcome types, the fraction of participants observed dead in the intervention group is equal to the fraction who would be killed plus the fraction who would die regardless. (By definition, those who would be saved and those who would live regardless would not contribute to the fraction observed dead in the intervention group.) By similar reasoning, the fraction of participants observed dead in the control group is equal to the fraction who would be saved plus the fraction who would die regardless. Because the randomization assures that the expected fraction who would die regardless is the same in the intervention and control

2

groups, the reduced form is equal to the difference between the fraction of participants who would be killed and the fraction of participants who would be saved.

Beyond the *difference* between the fraction of participants who would be killed and the fraction of participants who would be saved, it could also be useful to know the ratio of the two fractions and the absolute numbers of participants saved and killed. For example, in moral philosophy, the "trolley problem" and the related "transplant problem," involve whether it is ethical to kill one person to save five others, either by diverting a trolley or by reallocating organs (Foot, 1967; Thomson, 1985). Randomized experiments no doubt involve ethical questions that are difficult to answer, and it is perhaps even more difficult to answer those questions without the numbers of participants who would be saved and killed.

I develop estimators of the numbers of participants of each of the four potential outcome types: those who would be saved, those who would be killed, those who would die regardless, and those who would live regardless. I begin by developing a model of a randomized experiment. The model relies on the randomization itself and on deductive reasoning. Using the model, I derive an expression for the probability of the data that I observe as a function of the number of participants of each potential outcome type and the intended fraction of participants in the intervention group. I then propose two estimators of the number of participants of each potential outcome type. The first is a maximum likelihood estimator. It maximizes the likelihood of the number of participants of each potential outcome type given the data that I observe and the intended fraction of participants in the intervention group. The maximum likelihood estimator is not yet computationally tractable via a variety of approaches, so I develop a second estimator. The second estimator is a least squares estimator that minimizes a weighted function of the randomization error within and across potential outcome types, and it is computationally tractable.

I apply the least squares estimator to data from an important clinical trial, the Recombinant Human Activated Protein C Worldwide Evaluation of Severe Sepsis (PROWESS) clinical trial (Bernard et al., 2001). This trial randomized access to an intravenously-administered biologic drug referred to as recombinant human activated protein C, drotrecogin alfa activated, or Xigris (manufactured by Eli Lilly). Trial participants included 1690 patients with severe sepsis, a life-threatening condition. The reduced form for 28-day mortality showed that the intervention reduced mortality by 6 percentage points, which was sizable relative to the 31% 28-day mortality rate in the control group. On the basis of this result, the trial suspended enrollment, and the FDA expedited approval of the drug for patients with severe sepsis in 2001.

However, the FDA approval was controversial in part because an alternative reduced form within the trial showed that the intervention increased serious bleeding by 1.5 percentage points (Siegel, 2002; Warren et al., 2002). The increase in serious bleeding provides

3

a potential mechanism through which the intervention could kill participants. The article that presented the trial results claimed that "1 additional life would be saved for every 16 patients" randomized into the intervention group (Bernard et al., 2001). Presumably, the origin of this claim is that 1/16 is approximately 6 percentage points, the size of the reduced form. However, given the relationship between the reduced form and the number of patients saved and killed, it would be more accurate to say that 1 additional life would be saved *on net* for every 16 patients randomized into the intervention group. The reduced form indicates that the intervention saved 6 percentage points more participants than it killed, but the number of participants killed, if any, is not observable, and neither is the ratio of participants killed to saved. The model paves the way for me to estimate those statistics.

To apply the model, I need very limited aggregate data, which I take directly from the article that presented the trial results (Bernard et al., 2001). Specifically, the data include the number of participants who died in the intervention group, the number of participants who lived in the intervention group, the number of participants who died in the control group, and the number of participants who lived in the control group. I also take as given the intended fraction of participants in the intervention group, which was reported to be one half.

Within the PROWESS trial, I find that 57% of participants would live regardless, 18% would be saved, 12% would be killed, and 13% would die regardless. Putting these numbers together, the intervention would have no effect on 70% of participants, but among the remaining participants, it would kill two participants for every three it saved (12%/18%=2/3). I find that the total number of participants who would be killed by the intervention is 205. However, about half of those participants were were randomized into the control group, so they were not killed. The intermediate output of my estimator shows that 103 participants were killed within the trial. These results raise ethical questions about whether to expand access to the drug. It is notable that based on subsequent evidence, the manufacturer, Eli Lilly, voluntarily withdrew the drug from the market worldwide in 2011 (US Food and Drug Administration and others, 2011).

To assess the performance of my estimator, I perform a Monte Carlo exercise calibrated to the PROWESS trial. Because the experiment and the Monte Carlo exercise both involve randomization, the Monte Carlo exercise provides an intuitive illustration of the model. I begin by generating experiments of the same size as the PROWESS trial. In each experiment, I calibrate the true number of participants of each potential outcome type to my estimates from the trial, and I generate the data. I then apply my estimator to recover new estimates in each experiment. Across simulations, I find that the average bias in the estimated number of participants of each potential outcome type is fairly stable across potential outcome types at approximately 42 participants, which represents 2.5% of the full sample size. The RMSE

in the estimate of each potential outcome type is approximately 7.3% of the full sample size.

In the next section, I introduce the model, which I use to derive the probability of the observed data. I present the maximum likelihood estimator and the least squares estimator in Section 3. I present the empirical application to the PROWESS trial and the related Monte Carlo simulation in Sections 4 and 5. I conclude in Section 6.

2 Model

Consider a participant with potential outcome type i. For example, the participant could be of the type that would live regardless of the intervention. He does not know his potential outcome type. He does not even know the distribution of potential outcome types, so he is willing to be randomized into the intervention group or the control group.

At the start of the experiment, the experimenter has already determined the intended fraction of participants in the intervention group p. If $p = 1/2$, then the experimenter flips a fair coin to determine whether the participant will be randomly assigned to the intervention or control group. If $p \neq 1/2$, then the experimenter uses a different randomization method. Either way, the indicator for whether the participant is assigned to the intervention group is a random variable that is distributed according to a Bernoulli distribution with probability of success p.

Suppose that there are other participants of the same potential outcome type as the participant in question. Denote the total number of participants of potential outcome type i as $t(i)$. Randomization occurs independently for each participant with the same probability of assignment to the intervention group p. Therefore, the total number of participants of potential outcome type i who are randomly assigned to the intervention group is equal to the sum of independent and identically distributed Bernoulli random variables. The sum of independent and identically distributed Bernoulli random variables has a binomial distribution with parameters that represent the number of trials and the probability of success in each trial. Therefore, the total number of participants of potential outcome type i who are randomly assigned to the intervention group is distributed according to a binomial distribution with parameters $t(i)$ and p.

There are four possible potential outcome types. The matrix in Figure 1 includes a separate row for each potential outcome type i. The first row includes the $t(1)$ participants who would live regardless of the intervention ($Y = 0$ if $Z = 1$ and $Y = 0$ if $Z = 0$); the second row includes the $t(2)$ participants who would be saved by the intervention ($Y = 0$ if $Z = 1$ and $Y = 1$ if $Z = 0$); the third row includes the $t(3)$ participants who would be killed by the intervention ($Y = 1$ if $Z = 1$ and $Y = 0$ if $Z = 0$); and the fourth row includes the $t(4)$ participants who would die regardless of the intervention ($Y = 1$ if $Z = 1$ and $Y = 1$ if $Z = 0$).

Figure 1: Matrix that Relates Potential Outcome Types and Observed Outcome Groups

Potential Outcome Type i			Dead in intervention $Z = 1$ $Y = 1$ G(1)	Alive in intervention $Z = 1$ $Y = 0$ G(2)	Dead in control $Z = 0$ $Y = 1$ G(3)	Alive in control $Z = 0$ $Y = 0$ G(4)
				Observed Outcome Group j		
Would live regardless:	$Y = \begin{cases} 0 \text{ if } Z = 1 \\ 0 \text{ if } Z = 0 \end{cases}$	t(1)		N(1,2)		N(1,4)
Would be saved:	$Y = \begin{cases} 0 \text{ if } Z = 1 \\ 1 \text{ if } Z = 0 \end{cases}$	t(2)		N(2,2)	N(2,3)	
Would be killed:	$Y = \begin{cases} 1 \text{ if } Z = 1 \\ 0 \text{ if } Z = 0 \end{cases}$	t(3)	N(3,1)			N(3,4)
Would die regardless:	$Y = \begin{cases} 1 \text{ if } Z = 1 \\ 1 \text{ if } Z = 0 \end{cases}$	t(4)	N(4,1)		N(4,3)	

Note. Y represents mortality, and Z represents assignment to the intervention group. $N(i,j)$, the number of participants of potential outcome type i in observed outcome group j, must be equal to zero in all shaded cells. The intended fraction of participants in the intervention group is p.

Importantly, there is effectively a separate randomized experiment within each potential outcome outcome type i. Therefore, the total number of participants of any potential outcome type who are randomly assigned to the intervention group is independent of the total number of participants of any other potential outcome type who are randomly assigned to the intervention group. This independence proves useful in expressing the distribution of the data as a function of independent binomial random variables.

The data consist of the total number of participants observed in each of four possible observed outcome groups at the end of the experiment, after the intervention group has received access to the treatment. The matrix in Figure 1 includes a separate column for each observed outcome group j. The first column includes the $G(1)$ participants observed dead in the intervention group ($Z = 1$ and $Y = 1$); the second column includes the $G(2)$ participants observed alive in the intervention group ($Z = 1$ and $Y = 0$); the third column

includes the $G(3)$ participants observed dead in the control group ($Z = 0$ and $Y = 1$); and the fourth column includes the $G(4)$ participants observed alive in the control group ($Z = 0$ and $Y = 0$). Note that $G(j)$ is a random variable, which is why I denote it with a capital letter. The data consist of the vector of $G(1)$, $G(2)$, $G(3)$, and $G(4)$, which I represent with **G**, in bold to indicate that it is a vector. In what follows, I denote a realization of $G(i)$ with $g(i)$ and a realization of **G** with **g**. I also denote the vector of $t(1)$, $t(2)$, $t(3)$, and $t(4)$ with **t**.

The purpose of my model is to yield an expression for $P(\mathbf{G} = \mathbf{g} \mid \mathbf{t}, p)$, the probability of the data vector **G** in terms of the vector of potential outcome types **t** and the intended fraction of participants in the intervention group p. Toward that end, the matrix in Figure 1 relates the four potential outcome types to the four observed outcome groups. Each cell in the matrix includes the $N(i, j)$ participants with potential outcome type i in observed outcome group j. In what follows, I denote the matrix of all $N(i, j)$ with **N**, in bold to indicate that it is a matrix. The matrix **N** has 16 total cells.

However, 8 of the cells cannot have any participants in them. Consider the upper left cell, which includes participants of the potential outcome type $i = 1$ and observed outcome group $j = 1$. Participants of potential outcome type $i = 1$ would live regardless of assignment to the intervention group, and participants of observed outcome group $j = 1$ are observed dead in the intervention group. By definition, participants of the type who would live regardless of the assignment to the intervention group will not be observed dead in the intervention group, so it must be the case that $N(1, 1) = 0$. Similarly, $N(1, 3) = 0$ participants of the type $i = 1$ who would live regardless of assignment to the intervention group will not be observed dead in the control group $j = 3$. The logic for other cells proceeds similarly. I shade all 8 cells that cannot have any participants in them.

By deductive reasoning based on the shaded cells, I can express the distribution of the data as follows:

$$P(\mathbf{G} = \mathbf{g} \mid \mathbf{t}, p) = P\Big(G(1) = g(1), G(2) = g(2), G(3) = g(3), G(4) = g(4) \mid \mathbf{t}, p\Big) \quad (1)$$

$$= P\Big(N(3, 1) + N(4, 1) = g(1),$$

$$N(1, 2) + N(2, 2) = g(2),$$

$$N(2, 3) + N(4, 3) = g(3),$$

$$N(1, 4) + N(3, 4) = g(4) \mid \mathbf{t}, p\Big) \quad (2)$$

$$= P\Big(N(3, 1) + N(4, 1) = g(1),$$

$$N(1, 2) + N(2, 2) = g(2),$$

7

$$N(2,2) + N(4,1) = t(2) + t(4) - g(3),$$

$$N(1,2) + N(3,1) = t(1) + t(3) - g(4) \mid \mathbf{t}, p \Big). \tag{3}$$

I transition from (1) to (2) by expressing each $G(j)$ as the sum of the nonzero participant counts in column j. The resulting expression is in terms of the participant counts in all 8 cells that are not shaded in Figure 1. I simplify the expression further by recognizing that within any potential outcome type i, the number of participants in the control group is equal to $t(i)$ minus the number of participants in the intervention group. Therefore, I can express (3) in terms of only four random variables: $N(1,2)$, $N(2,2)$, $N(3,1)$, and $N(4,1)$. These random variables represent the realized number of participants assigned to the intervention group in each of the four potential outcome types.

From above, we know that the number of participants assigned to the intervention group in each potential outcome type is a binomial random variable with parameters $t(i)$ and p. Furthermore, we know that the number of participants assigned to the intervention group is independent for each type. Therefore, (3) is a convolution of four independent probabilities with known distributions. I aim to express (3) as the product of four independent probabilities with known distributions.

Using the approach to deconvolution for discrete random variables, I express (3) as the sum of the probability of each possible realization ℓ of $N(1,2)$. I then rearrange terms to obtain an expression in terms of the joint probability of $N(1,2)$, $N(2,2)$, $N(3,1)$, and $N(4,1)$:

$$P(\mathbf{G} = \mathbf{g} \mid \mathbf{t}, p) = \sum_{\ell=0}^{t(1)} P\Big(N(1,2) = \ell,$$

$$N(3,1) + N(4,1) = g(1),$$
$$\ell + N(2,2) = g(2),$$
$$N(2,2) + N(4,1) = t(2) + t(4) - g(3),$$
$$\ell + N(3,4) = t(1) + t(3) - g(4) \mid \mathbf{t}, p \Big)$$

$$= \sum_{\ell=0}^{t(1)} P\Big(N(1,2) = \ell,$$

$$N(2,2) = g(2) - \ell,$$
$$N(3,1) = t(1) + t(3) - g(4) - \ell,$$
$$N(4,1) = g(1) + g(4) + \ell - t(1) - t(3),$$

8

$$\sum_{j=1}^{4} g(j) = \sum_{i=1}^{4} t(i) \mid \mathbf{t}, p\bigg).$$

Since $N(1,2)$, $N(2,2)$, $N(3,1)$, and $N(4,1)$ are independently distributed binomial random variables, it is possible to express the distribution of the data as follows:

$$\begin{aligned}
P(\mathbf{G} = \mathbf{g} \mid \mathbf{t}, p) = \sum_{\ell=0}^{t(1)} &P\Big(N(1,2) = \ell \mid \mathbf{t}, p\Big) \\
&\times P\Big(N(2,2) = g(2) - \ell \mid \mathbf{t}, p\Big) \\
&\times P\Big(N(3,1) = t(1) + t(3) - g(4) - \ell \mid \mathbf{t}, p\Big) \\
&\times P\Big(N(4,1) = g(1) + g(4) + \ell - t(1) - t(3) \mid \mathbf{t}, p\Big) \\
&\times \mathbf{1}\left\{\sum_{j=1}^{4} g(j) = \sum_{i=1}^{4} t(i)\right\} \\
= \sum_{\ell=0}^{t(1)} &\mathrm{binom}\Big(\ell, t(1), p\Big) \\
&\times \mathrm{binom}\Big(g(2) - \ell, t(2), p\Big) \\
&\times \mathrm{binom}\Big(t(1) + t(3) - g(4) - \ell, t(3), p\Big) \\
&\times \mathrm{binom}\Big(g(1) + g(4) + \ell - t(1) - t(3), t(4), p\Big) \\
&\times \mathbf{1}\left\{\sum_{j=1}^{4} g(j) = \sum_{i=1}^{4} t(i)\right\}
\end{aligned}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and $\mathrm{binom}(\cdot)$ is the binomial probability mass function,

$$\mathrm{binom}(k, r, p) = P(K = k \mid r, p) = \binom{r}{k} p^k (1-p)^{r-k}.$$

3 Estimation

3.1 Maximum Likelihood Estimator

Using the probability of the realized data vector $\mathbf{g}$ in terms of the vector of potential outcome types $\mathbf{t}$ and the intended fraction of participants in the intervention group p, I can express the likelihood of the vector of potential outcome types $\mathbf{t}$ in terms of the realized data vector $\mathbf{g}$ and the intended fraction of participants in the intervention group p as follows:

$$\mathcal{L}(\mathbf{t} \mid \mathbf{g}, p) = P(\mathbf{G} = \mathbf{g} \mid \mathbf{t}, p)$$

9

I use the likelihood to specify the following maximum likelihood estimator

$$\max_{\mathbf{t} \in \mathbb{N}_0^4} \; \mathcal{L}(\mathbf{t} \mid \mathbf{g}, p)$$

where $\mathbf{t} \in \mathbb{N}_0^4$ indicates that the four elements of the vector $\mathbf{t}$ must belong to the set that includes the natural numbers and zero, which follows because they represent the counts of participants of each potential outcome type i.

As specified, the maximum likelihood estimator is a nonlinear integer programming problem. It is possible that an analytical solution to this problem exists, but I have not be able to attain it using the software package Mathematica (Wolfram Research, Inc., 2018).[1] I have also not been able to attain a numerical solution using Mathematica.[2] It is possible that this problem could be solved numerically with another software package via a grid search, but I have not been able to find a solution through such an approach because the size of the grid is large for experiments of reasonable size, and evaluation of the objective function produces numbers that are the same within machine precision. Software packages such as BARON (Sahinidis, 2018), can solve some types of nonlinear integer programming problems numerically when grid search algorithms are intractable. However, BARON cannot solve this problem directly because it does not allow binomial coefficient terms to be specified within the objective function, nor does it allow a choice variable to appear as the upper limit of summation.[3] I see the application of computational advances to this nonlinear programming problem as a promising avenue for future work.

3.2 Least Squares Estimator

Since the maximum likelihood estimator is not yet computationally tractable via a variety of approaches, I propose an alternative least squares estimator that is. Intuitively, the maximum likelihood estimator maximizes a function that depends on the known functional form of the randomization error within the experiment. The least squares estimator minimizes a function of the randomization error within the experiment using moments of its known functional form.

To be precise, define $\varepsilon(i)$, the randomization error within each potential outcome type

---

[1] Mathematica's Maximize command is unable to maximize this expression and simply returns the input. Using real interpolations of the factorial function, I can obtain partial derivatives of the likelihood function, but Mathematica's Solve command cannot find a critical point, returning the error "Solve::nsmet: This system cannot be solved with the methods available to Solve."

[2] Mathematica's NSolve function hangs indefinitely, while FindRoot results in errors due to limitations of machine-precision numbers.

[3] BARON limits the structure of the objective and constraint functions. These functions must be expressed in terms of a constant number of algebraic operations. While binomial coefficient terms can be specified as a string of products, the number of terms in the product depends on the input values. Similarly, varying the upper limit of summation results in a varying number of summation terms.

i, as the difference between the actual and expected numbers of participants of type i randomized into the intervention group, in terms of the intended fraction of participants in the intervention group p, the number of type i participants randomized into the intervention group $N(i,1) + N(i,2)$, and the total number of type i participants $t(i)$:

$$\varepsilon(i) = N(i,1) + N(i,2) - p * t(i).$$

From the model, we know that $N(i,1) + N(i,2)$ is distributed according to a binomial distribution with parameters $t(i)$ and p. The mean of a binomial random variable is equal to the product of the parameters of the binomial distribution. Therefore, it follows that the mean of the randomization error $\varepsilon(i)$ within each potential outcome type i is zero: $E[\varepsilon(i)] = p * t(i) - p * t(i) = 0$.

Furthermore, we can characterize the variance of the randomization error within each potential outcome type i. The variance of a binomial random variable is equal to the product of the probability of success, the probability of failure, and the number of trials. Subtracting a constant from a binomial random variable does not change its variance. Therefore, it follows that $Var[\varepsilon(i)] = p(1-p) * t(i)$. Note that the variance of the randomization error increases as the number of participants increases. A randomization error of 10 would be unlikely in a sample of 10, but it would be much more likely in a sample of 1000.

Just as is it possible to characterize the mean and variance of the randomization error within each potential outcome type, it is also possible to characterize the mean and variance of the randomization error in any sample that combines all participants from two or more potential outcome types, including the experiment as a whole. Let $I \subseteq \{1,2,3,4\}$ represent any subset of the potential outcome types. Using this notation, we can express the the randomization error in any sample formed by combining all participants from every potential outcome type included in I as follows:

$$\varepsilon(I) = \sum_{i \in I} [N(i,1) + N(i,2)] - p \sum_{i \in I} t(i). \tag{4}$$

It is trivial show that $\varepsilon(I)$ has mean zero and variance equal to $p(1-p) \sum_{i \in I} t(i)$, which is proportional to the total number of participants $\sum_{i \in I} t(i)$.

Using the mean and the variance of the randomization error in each subset of potential outcome types I, I construct an objective function $\mathcal{S}$. This objective function represents the sum of the squared randomization error, weighted by the inverse of the variance of the randomization error, across all subsets of potential outcome types I within the experiment:

$$\mathcal{S}(\mathbf{N} \mid p) = \sum_{I} \frac{1}{p(1-p) \sum_{i \in I} t(i)} \varepsilon(I)^2$$

$$= \sum_I \frac{1}{p(1-p)\sum_{i\in I}\{N(i,1) + N(i,2) + N(i,3) + N(i,4)\}}$$

$$\times \left\{ \sum_{i\in I}[N(i,1) + N(i,2)] - p\sum_{i\in I}t(i) \right\}^2,$$

where I have shown that $\mathcal{S}$ is a function of the elements $N(i,j)$ of the matrix $\mathbf{N}$ by substituting for $\varepsilon(I)$ via (4) and substituting by for $t(i) = N(i,1) + N(i,2) + N(i,3) + N(i,4)$. Intuitively, because the variance of the randomization error should be larger in larger subsamples, the objective function imposes a larger penalty for randomization error in smaller subsamples.

The least squares estimator minimizes the objective function $\mathcal{S}$ with respect to $\mathbf{N}$, subject to constraints that impose that $\mathbf{N}$ must give rise to the observed data vector $\mathbf{G}$ and that some elements of $\mathbf{N}$ must be equal to zero by deductive reasoning based on the shaded cells of the matrix in Figure 1:

$$\min_{\mathbf{N}\in\mathbb{N}_0^{4\times 4}} S(\mathbf{N} \mid p)$$

$$\text{s.t. } \sum_{i=1}^{4} N(i,j) = g(j) \text{ for } j = 1,2,3,4$$

$$N(1,1) = N(1,3) = N(2,1) = N(2,4) = N(3,2) = N(3,3) = N(4,2) = N(4,4) = 0,$$

where $\mathbf{N} \in \mathbb{N}_0^{4\times 4}$ indicates that each element $N(i,j)$ of the 4 by 4 matrix $\mathbf{N}$ must belong to the set that includes the natural numbers and zero. I denote the estimate of each $N(i,j)$ with $N(\hat{i},j)$, and I construct the least squares estimate of each potential outcome type $\hat{t}(i)$ as follows:

$$\hat{t}(i) = \sum_{j=1}^{4} \hat{N}(i,j) \qquad \forall i = 1,2,3,4.$$

Unlike the maximum likelihood estimator, the least squares estimator involves a nonlinear programming problem of a form that can be estimated numerically in BARON.

4 Empirical Application

4.1 The PROWESS Clinical Trial

To apply my model to the PROWESS clinical trial, I use aggregated data reported in Bernard et al. (2001), which allow me to replicate the main reduced form result from Bernard et al. (2001) exactly. The main reduced form result is based on the comparison of 28-day mortality ($Y = 1$ if dead, $Y = 0$ if alive) for 850 participants assigned to the intervention group and 840 participants assigned to the control group ($Z = 1$ if intervention, $Z = 0$ if control). Bernard

et al. (2001) report that participants were randomized into the intervention and control groups in a 1:1 manner, so I set the intended fraction of participants in the intervention group to $p = 1/2$.

By day 28, 210 participants in the intervention group had died and 259 participants in the control group had died. Therefore, as I report across the columns of the matrix in Figure 2, $g(1) = 210$ participants were dead in the intervention group, $g(2) = 640$ participants were alive in the intervention group, $g(3) = 259$ participants were dead in the control group, and $g(4) = 581$ participants were alive in the control group. The reduced form based on these statistics shows that assignment to the intervention group decreased mortality by 6 percentage points (0.06 is approximately equal to 210/850 - 259/840), which is statistically different from 0 at the 0.5% level. An alternative reduced form reported in Bernard et al. (2001) shows that assignment to the intervention group increased the incidence of serious bleeding by 1.5 percentage points, which is statistically different from 0 at the 6% level, suggesting a potential mechanism through which participants could be killed within the trial.

4.2 Results

I report the results that I obtained by applying the least squares estimator to the PROWESS trial in the rows of the matrix in Figure 2. I obtained the estimates by using the BARON optimization package (Sahinidis, 2018) within MATLAB (The Mathworks, Inc., 2016). The estimates show that $t(1) = 964$ participants would live regardless, $t(2) = 308$ participants would be saved, $t(3) = 205$ participants would be killed, and $t(4) = 213$ participants would die regardless.

As a fraction of the full sample of 1690 participants, 57% would live regardless, 18% would be saved, 12% would be killed, and 13% would die regardless. Note that the absolute difference between the number of participants saved and killed is 6%, which is mechanically equal to the reduced form estimate. However, the least squares estimates provide context for the reduced form estimate by showing that 70% of participants would not be affected by assignment to the intervention and that 2 participants would be killed for every 3 participants saved (2/3=12%/18%). Because of the random assignment, only about half of the patients who would be killed were assigned to the intervention group. One element of the intermediate output of my least squares estimator, $n(3, 1)$, reported in the first column of the third row of the matrix in Figure 2, shows that 103 trial participants were killed.

To assess the statistical significance of these estimates, I report standard errors that I obtained as the standard deviation of the estimates that I obtained via 1000 bootstrap iterations. The standard errors are very similar across the potential outcome types, so statistical significance tends to increase with the number of participants of each potential outcome type. The number of participants who would live regardless is the largest ($t(1) = 964$), and

Figure 2: Matrix that Relates Potential Outcome Types and Observed Outcome Groups Estimates from the PROWESS Trial

			Dead in intervention	Alive in intervention	Dead in control	Alive in control
Observed Outcome Group j			$Z = 1$ $Y = 1$	$Z = 1$ $Y = 0$	$Z = 0$ $Y = 1$	$Z = 0$ $Y = 0$
			g(1)	g(2)	g(3)	g(4)
			210	640	259	581
Potential Outcome Type i			12%	38%	15%	34%
Would live regardless: $Y = \begin{cases} 0 \text{ if } Z = 1 \\ 0 \text{ if } Z = 0 \end{cases}$	t(1)	964 (114) 57%	■	485 (59) 29%	■	479 (58) 28%
Would be saved: $Y = \begin{cases} 0 \text{ if } Z = 1 \\ 1 \text{ if } Z = 0 \end{cases}$	t(2)	308 (114) 18%	■	155 (57) 9%	153 (57) 9%	■
Would be killed: $Y = \begin{cases} 1 \text{ if } Z = 1 \\ 0 \text{ if } Z = 0 \end{cases}$	t(3)	205 (113) 12%	103 (57) 6%	■	■	102 (56) 6%
Would die regardless: $Y = \begin{cases} 1 \text{ if } Z = 1 \\ 1 \text{ if } Z = 0 \end{cases}$	t(4)	213 (113) 13%	107 (57) 6%	■	106 (56) 6%	■

Note. Bootstrapped standard errors are reported in parentheses. Percentages represent percentages of the full sample. They do not necessarily sum to 100% due to rounding. Y represents 28-day mortality, and Z represents assignment to the intervention group in the PROWESS trial. $N(i, j)$, the number of participants of potential outcome type i in observed outcome group j, must be equal to zero in all shaded cells. The observed outcome group vector **g** and the intended fraction of participants in the intervention group $p = 1/2$ are taken from the PROWESS trial. The values of the matrix **N** and the potential outcome type vector **t** are estimates obtained via the least squares estimator, estimated using the BARON optimization package within MATLAB.

it is statistically different from zero at conventional levels. The number of participants who would be saved ($t(2) = 308$) is also statistically different from zero at conventional levels.

5 Monte Carlo

5.1 Design

To examine the performance of the least squares estimator, I perform a Monte Carlo simulation. In each simulated experiment m, I set the true vector of the number of participants of each potential outcome type **t** to the estimated vector from the PROWESS trial. Using the intended fraction of participants in the intervention group $p = 1/2$, I randomly assign each participant to the intervention or control group. I use each participant's potential outcome

type i to determine each participant's observed outcome group j. For example, if I assign a participant who would live regardless ($i = 1$) to the intervention group, then I recognize that the participant will live and thus be observed in outcome group $j = 2$. Aggregating across all participants within a simulated experiment, I determine the observed outcome group vector $\mathbf{g}$. Using the observed outcome group vector $\mathbf{g}$ and the intended fraction of participants in the intervention group $p = 1/2$, I apply the least squares estimator and store my estimates of the potential outcome type vector $\mathbf{t}$. I run a total of $M = 1000$ simulated experiments. Across all experiments, I construct the mean bias and root mean square error (RMSE) for each element of the potential outcome type vector $\mathbf{t}$.

5.2 Results

I report the results of the Monte Carlo simulation in Table 1. The results show that the mean bias and RMSE do not vary much across potential outcome types. For each potential outcome type i, the mean bias in the total number of participants $t(i)$ is approximately 42 participants, which represents 2.5% of the full sample of 1690 participants. The RMSE is approximately 124 participants for each type i, which represents 7.3% of the full sample of 1690 participants.

Table 1: Monte Carlo Simulation Results

Potential Outcome Type i			True	Bias	RMSE
$t(1)$	Would live regardless:	$Y = \begin{cases} 0 \text{ if } Z = 0 \\ 0 \text{ if } Z = 1 \end{cases}$	964 (57.0%)	-41.50 (-2.5%)	124.88 (7.4%)
$t(2)$	Would be saved:	$Y = \begin{cases} 1 \text{ if } Z = 0 \\ 0 \text{ if } Z = 1 \end{cases}$	308 (18.2%)	41.50 (2.5%)	126.46 (7.5%)
$t(3)$	Would be killed:	$Y = \begin{cases} 0 \text{ if } Z = 0 \\ 1 \text{ if } Z = 1 \end{cases}$	205 (12.1%)	41.55 (2.5%)	123.07 (7.3%)
$t(4)$	Would die regardless:	$Y = \begin{cases} 1 \text{ if } Z = 0 \\ 1 \text{ if } Z = 1 \end{cases}$	213 (12.6%)	-41.55 (-2.5%)	123.80 (7.3%)

Note. The Monte Carlo simulation generates $M = 1000$ simulated experiments in which the true potential outcome type vector $\mathbf{t}$ is equal to the estimated potential outcome type vector from the PROWESS trial. In each simulation, the observed outcome group vector $\mathbf{g}$ is calculated via the model, using the intended fraction of participants $p = 1/2$ from the PROWESS trial. Within each simulated experiment, $\mathbf{g}$ and $p = 1/2$ are the inputs used to estimate the potential outcome type vector $\mathbf{t}$ with the least squares estimator, via the BARON optimization package within MATLAB. Letting $t(i)$ be the number of participants of true potential outcome group i and $\hat{t}_m(i)$ be the estimate of number of participants of true potential outcome group i in simulation m, mean bias is $\frac{1}{M} \sum_m (\hat{t}_m(i) - t(i))$, and RMSE is $\left(\frac{1}{M} \sum_m (\hat{t}_m(i) - t(i))^2 \right)^{1/2}$. Percentages represent percentages of the full sample.

6   Conclusion

In this paper, I develop a model of a randomized experiment. Although I have focused on a randomized experiment, the model could also be adapted for application to any natural experiment or instrumental variable setting. When applied to the PROWESS clinical trial, the model and the associated least squares estimator that I develop allow me to estimate the number of participants who would live regardless, the number of participants who would be saved, the number of participants who would be killed, and the number of participants who would die regardless. My estimates show that the intervention within the PROWESS trial killed two participants for every three it saved. These estimates have ethical implications.

References

Bernard, G. R., J.-L. Vincent, P.-F. Laterre, S. P. LaRosa, J.-F. Dhainaut, A. Lopez-Rodriguez, J. S. Steingrub, G. E. Garber, J. D. Helterbrand, E. W. Ely, and C. J. Fisher (2001). Efficacy and safety of recombinant human activated protein c for severe sepsis. *New England Journal of Medicine 344*(10), 699–709. PMID: 11236773.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review 5*, 5–15.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association 81*(396), 945–960.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral statistics 2*(1), 1–26.

Sahinidis, N. V. (2018). *BARON 18.8.23: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual.

Siegel, J. P. (2002). Assessing the use of activated protein c in the treatment of severe sepsis. *The New England journal of medicine 347*(13), 1030–1034.

The Mathworks, Inc. (2016). Matlab, Version r2016a. Natick, Massachusetts, United States.

Thomson, J. J. (1985). The trolley problem. *Yale Law Journal 94*(1395), 1395–1415.

US Food and Drug Administration and others (2011). FDA drug safety communication: voluntary market withdrawal of xigris due to failure to show a survival benefit. *US Food and Drug Administration, Washington, DC*.

Warren, H. S., A. F. Suffredini, P. Q. Eichacker, and R. S. Munford (2002). Risks and benefits of activated protein c treatment for severe sepsis. *The New England journal of medicine 347*(13), 1027–1030.

Wolfram Research, Inc. (2018). Mathematica, Version 11.3. Champaign, IL.