Received: 10 September 2020 | Revised: 10 August 2021 | Accepted: 20 August 2021

DOI: 10.1111/biom.13554

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Causal inference with outcomes truncated by death in multiarm studies

Shanshan Luo[1] | **Wei Li**[2] | **Yangbo He**[1]

[1] School of Mathematical Sciences, Peking University, Beijing, China

[2] Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China

Correspondence
Wei Li, Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing 100872, China.
Email: weilistat@ruc.edu.cn

Funding information
NSFC, Grant/Award Numbers: 11671020, 11971040; Huawei Technologies; The Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China

Abstract

It is challenging to evaluate causal effects when the outcomes of interest suffer from truncation-by-death in many clinical studies; that is, outcomes cannot be observed if patients die before the time of measurement. To address this problem, it is common to consider average treatment effects by principal stratification, for which, the identifiability results and estimation methods with a binary treatment have been established in previous literature. However, in multiarm studies with more than two treatment options, estimation of causal effects becomes more complicated and requires additional techniques. In this article, we consider identification, estimation, and bounds of causal effects with multivalued ordinal treatments and the outcomes subject to truncation-by-death. We define causal parameters of interest in this setting and show that they are identifiable either using some auxiliary variable or based on linear model assumption. We then propose a semiparametric method for estimating the causal parameters and derive their asymptotic results. When the identification conditions are invalid, we derive sharp bounds of the causal effects by use of covariates adjustment. Simulation studies show good performance of the proposed estimator. We use the estimator to analyze the effects of a four-level chronic toxin on fetal developmental outcomes such as birth weight in rats and mice, with data from a developmental toxicity trial conducted by the National Toxicology Program. Data analyses demonstrate that a high dose of the toxin significantly reduces the weights of pups.

KEYWORDS
bounds, developmental toxicity, estimation, multiarm study, truncation-by-death

1 | INTRODUCTION

In many multiarm chronic toxicity studies, it is common that the interested nonmortality outcomes are truncated by death during a long-term experiment. For instance, in our motivating example from a developmental toxicity study conducted by the National Toxicology Program (NTP; Price et al., 1985; Elliott et al., 2006; NTP, 2017), pup mice and rats were randomly exposed to one of four ordinal toxin doses by whole body inhalation. Laboratory scientists were interested in evaluating relative effects of toxin doses on the weights of pups at the end of 2 years after the receipt of treatment. However, during this long-term study, some pups died before their final weights were measured; that is, their outcomes are truncated by death, and hence the outcomes of these dead individuals are not well defined. Direct comparison of pups that survived at high doses of the toxin with those

that survived at low doses may lead to selection bias, because some unhealthy pups may be included in the latter group, and they would die if exposed to higher levels of toxins.

Methods for dealing with such truncation-by-death data rely heavily on the principal stratification approach proposed by Frangakis and Rubin (2002). The main parameter of interest in these studies is the principal causal effect in the subpopulation of pups who would survive irrespective of toxin levels, which is also termed (always-) survivor average causal effect (Rubin et al., 2006). Although some criticisms may question about its interpretations (e.g., Pearl, 2011), the principal stratification framework has been widely adopted as a tool to handle truncation-by-death problems. One body of research considers the identifiability of the principal causal effect under various assumptions (Zhang et al., 2009; Ding et al., 2011; Tchetgen Tchetgen, 2014; Wang and Richardson, 2017). When these assumptions are invalid, an alternative strand of work focuses on estimation of bounds or conducts sensitivity analysis (Gilbert et al., 2003; Hayden et al., 2005; Shepherd et al., 2006; Lee et al., 2010; Chiba and VanderWeele, 2011; Ding and Lu, 2017).

However, most of previous work is devoted to studies with a binary treatment, and there has not been much discussion on truncation-by-death problems with multivalued ordinal treatments. From a practical point of view, it is necessary to develop approaches to handle this setting, because multiarm trials are fairly common in clinical studies, for example, a four-level treatment in our motivating example. Before our work, Frangakis et al. (2004) imposed the exclusion restriction assumption (Angrist et al., 1996) to guarantee identification of the principal causal effects. Some other scholars did not consider identification issues. For example, Elliott et al. (2006) proposed a hierarchical Bayesian approach to estimate the principal causal effects. A more recent work by Wang et al. (2017) developed a hypothesis testing method to detect nonnull principal causal effects for a binary outcome.

In this article, we consider identification, estimation, and bounds of principal causal effects in multiarm studies. Different from Frangakis et al. (2004), our identification assumptions allow the treatment to directly affect the outcome. We provide two sets of identification assumptions, in which the observed common causes of the treatment, survival status, and the outcome are allowed. This is more feasible in observational studies. We then propose a semiparametric method for estimating the principal causal effects and derive asymptotic results for the estimator. When the identification assumptions are invalid, we provide sharp bounds of the principal causal effects with available covariates and perform sensitivity analysis in our real data example.

The remainder of this paper is organized as follows. Section 2 introduces notation, definitions, and assumptions. In Section 3, we develop the identifiability results, estimation, and inference procedures for the principal causal effects. We also derive their sharp bounds when the identifiability conditions are violated. We illustrate the proposed approach via simulation studies and a real data set from the NTP experiment in Sections 4 and 5, respectively. All proofs and computational details are provided in the Supplementary Material.

2 | NOTATION, DEFINITION, AND ASSUMPTIONS

Assume that there are n individuals who are independently sampled from an infinite superpopulation of interest. We define the observed and potential random variables for individual i ($i = 1, \dots, n$). Notation for these variables suppress index i for simplicity. Let Z denote an ordinal treatment assignment with m ($m \geqslant 2$) levels: $Z \in \{1, \dots, m\}$, and let $\boldsymbol{X}$ denote a vector of covariates observed at baseline. Let S denote the survival status with 1 indicating survival and 0 for death. Let Y denote the nonmortality outcome of interest. In our example, the outcome Y denotes the logarithmic weight gain of a pup within 2 years. We use the potential outcomes framework and make the stable unit treatment value assumption; that is, there is only one version of potential outcomes and there is no interference between units (Rubin, 1990). Let $S(z)$ and $Y(z)$ denote the potential survival indicator and potential outcome that would be observed under treatment $Z = z$. Here, $Y(z)$ is well defined only if $S(z) = 1$. The observed values S and Y are deterministic functions of the treatment assignment and their respective potential values: $S = S(Z) = \sum_{z=1}^{m} I(Z = z)S(z)$ and $Y = Y(Z) = \sum_{z=1}^{m} I(Z = z)Y(z)$. Based on the potential survival status, we define the basic principal stratum as $G = \{S(1), \dots, S(m)\}$. Following Wang et al. (2017), we use the letter L to denote $S(z) = 1$ (meaning "live") and the letter D to denote $S(z) = 0$ (meaning "die"). Then G can be rewritten as a string consisting of letters L and D. For instance, in our example with a four-level toxic treatment, $G = LDDD$ indicates that a pup would survive at the control level 1 but would die at the toxicity level 2, 3, or 4. We introduce the basic assumptions which are commonly used in previous literature.

Assumption 1. $\{S(1), \dots, S(m), Y(1), \dots, Y(m)\} \perp\!\!\!\perp Z \mid \boldsymbol{X}$.

TABLE 1 Data structure under the monotonicity assumption

	Ordinal treatment and potential survival status				
Latent principal stratum	$Z=1$	$Z=2$	$\cdots$	$Z=m-1$	$Z=m$
$G=0$	$S(1)=0$	$S(2)=0$	$\cdots$	$S(m-1)=0$	$S(m)=0$
$G=1$	$S(1)=1$	$S(2)=0$	$\cdots$	$S(m-1)=0$	$S(m)=0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$G=m-1$	$S(1)=1$	$S(2)=1$	$\cdots$	$S(m-1)=1$	$S(m)=0$
$G=m$	$S(1)=1$	$S(2)=1$	$\cdots$	$S(m-1)=1$	$S(m)=1$
	Mixture of principal stratum				
Observed subgroup	$G=0$	$G=1$	$\cdots$	$G=m-1$	$G=m$
$Z=1,S=1$	$*$	$\surd$	$\cdots$	$\surd$	$\surd$
$Z=2,S=1$	$*$	$*$	$\cdots$	$\surd$	$\surd$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Z=m-1,S=1$	$*$	$*$	$\cdots$	$\surd$	$\surd$
$Z=m,S=1$	$*$	$*$	$\cdots$	$*$	$\surd$

Note: $\surd$: existence of the corresponding principal stratum. $*$: nonexistence of the corresponding principal stratum.

Assumption 1 means that the observed treatment assignment is independent of all potential survival status and potential outcomes conditional on observed covariates. This allows for possible observed confounders but precludes unobserved ones between treatment and survival status or the outcome, as the classical ignorable treatment assignment assumption does (Imbens and Rubin, 2015). Because the treatment assignment is randomized in our example, Assumption 1 is automatically satisfied.

Assumption 2. For each individual i, $S_i(z_1) \geqslant S_i(z_2)$ for all $z_1 \leqslant z_2$, where $z_1, z_2 \in \{1, \ldots, m\}$ are treatment levels.

Assumption 2 is a generalization of the monotonicity assumption for a binary treatment assignment imposed in Angrist et al. (1996). This monotonicity assumption in our example formally reflects that if a pup would not survive at some toxicity level, then the pup would also not survive at a higher toxicity level.

Under monotonicity assumption, each principal stratum G takes the form $L \cdots LD \cdots D$. To compress the notation, we write G as $L^k D^{m-k}$, or simply as label k for $k = 0, \ldots, m$. Each value of G is equivalent to the subject-specific threshold (maximum) level that the treatment assignment can be, above which the pup would not survive. In other words, the principal stratum G roughly represents the hidden physical condition of pups, and pups that are in good health are more resistant to toxicity. We present the ordinal structure of G and its relationship with the observed data in Table 1. For example, the subpopulation $\{i : G_i = 0\}$ includes pups that die at all treatment levels, and $\{i : G_i = k\}$ represents pups that survive only at treatment level $Z = z$ for $1 \leqslant z \leqslant k$, where $k = 1, \ldots, m$.

Analogous to the survivor average causal effect of a binary treatment, we can also define our causal parameters of interest as potential outcome contrasts within some basic principal stratum. Let $\mu_g(z) = E\{Y(z) \mid G = g\}$ denote the mean potential outcome under treatment assignment $Z = z$ within principal stratum $G = g$. As discussed earlier, for members of principal stratum $G = g$, $S(z) = 1$ if and only if the treatment level z satisfies $1 \leqslant z \leqslant g$. We thus define the pairwise causal estimands of interest as follows:

$$\Delta_g(z_1, z_2) = \mu_g(z_1) - \mu_g(z_2) \qquad \text{for } g \geqslant z_1 \geqslant z_2 \text{ and } g \geqslant 2. \tag{1}$$

Therefore, if $\mu_g(z)$ is identifiable for any $z \leqslant g$, then $\Delta_g(z_1, z_2)$ is also identifiable. In what follows, we focus on identification and estimation of $\mu_g(z)$ for $g \geqslant z$.

Let $\pi_k(\boldsymbol{x}) = \text{pr}(G = k \mid \boldsymbol{X} = \boldsymbol{x})$ denote the probability of principal stratum $G = k$ conditional on covariates $\boldsymbol{X} = \boldsymbol{x}$ for $k = 0, 1, \ldots, m$. Let $\mu(z, k, \boldsymbol{x}) = E(Y \mid Z = z, G = k, \boldsymbol{X} = \boldsymbol{x})$ denote the conditional mean response given $(Z = z, G = k, \boldsymbol{X} = \boldsymbol{x})$ for each $k \geqslant z$. Then $\mu_g(z)$ can be expressed as follows (see the Supplementary Material for proof):

$$\mu_g(z) = \pi_g^{-1} E\big\{\pi_g(\boldsymbol{X})\mu(z, g, \boldsymbol{X})\big\}, \tag{2}$$

where $\pi_g = E\{\pi_g(\boldsymbol{X})\}$. Under Assumptions 1 and 2, $\pi_k(\boldsymbol{x})$ is identifiable for each k. The intuition is as follows. Table 1 shows that for any treatment level $Z = z$ $(z = 1, \ldots, m)$, the observed subgroup $(Z = z, S = 1)$ is a mixture of the latent subgroups $(G = z), \ldots, (G = m)$; that is, the event

$\{Z = z, S = 1\}$ is equivalent to $\cup_{k=z}^{m}\{Z = z, G = k\}$. Hence, by Assumption 1,

$$\begin{aligned}\text{pr}(S = 1 \mid Z = z, \boldsymbol{X} = \boldsymbol{x}) &= \textstyle\sum_{k=z}^{m} \text{pr}(G = k \mid Z = z, \boldsymbol{X} \\ &= \boldsymbol{x}) = \textstyle\sum_{k=z}^{m} \pi_k(\boldsymbol{x})\end{aligned} \tag{3}$$

This implies that

$$\begin{aligned}\pi_k(\boldsymbol{x}) &= \text{pr}(S = 1 \mid Z = k, \boldsymbol{X} = \boldsymbol{x}) \\ -\text{pr}(S = 1 \mid Z = k + 1, \boldsymbol{X} = \boldsymbol{x}) &\ (1 \leqslant k \leqslant m - 1), \\ \pi_m(\boldsymbol{x}) &= \text{pr}(S = 1 \mid Z = m, \boldsymbol{X} = \boldsymbol{x}), \\ \pi_0(\boldsymbol{x}) &= 1 - \textstyle\sum_{k=1}^{m} \pi_k(\boldsymbol{x}).\end{aligned} \tag{4}$$

Thus, according to (2) and (4), the identifiability of $\mu_g(z)$ further requires identifiability of $\mu(z, g, \boldsymbol{x})$ for each $\boldsymbol{x}$.

3 | IDENTIFICATION, ESTIMATION, INFERENCE, AND BOUNDS

3.1 | Identification

In this subsection, we discuss further assumptions to identify $\mu_g(z)$. We consider two different assumptions: one is based on an auxiliary variable that is similar to an instrumental variable for the latent principal stratum, and the other relies on the specification of a linear model. Under the first assumption, $\mu_g(z)$ is nonparametrically identifiable. If we remove this assumption, then $\mu_g(z)$ can also be identifiable in a linear model setting.

For the first case, we assume that the covariates $\boldsymbol{X}$ can be written as $(A, \boldsymbol{C}^{\mathrm{T}})^{\mathrm{T}}$ such that the scalar covariate A affects the outcome Y only through its association with G conditional on treatment assignment Z and the remaining covariates $\boldsymbol{C}$. For convenience, we may use the notations $\boldsymbol{X}$ and $(A, \boldsymbol{C}^{\mathrm{T}})^{\mathrm{T}}$ interchangeably below.

Assumption 3. $Y \perp\!\!\!\perp A \mid (Z, G, \boldsymbol{C})$.

Assumption 3 precludes the direct effect of A on Y, which is similar to the exclusion restriction assumption in the instrumental variable analysis (Angrist et al., 1996). The main difficulty with the identifiability problem is that the principal stratum G is a latent variable. Assumption 3 makes it possible to identify $\mu_g(z)$ since the observed covariate A can be seen as a substitution variable for G. Assumption 3 is adapted from Ding et al. (2011), and we relax theirs by incorporating baseline covariates $\boldsymbol{C}$ in this assumption. Similar assumptions for a binary treatment can be found in Wang et al. (2017). Figure 1 gives a causal diagram illustrating Assumption 3. In our motivating example, we take A as the baseline logarithmic weight of each pup. To alleviate the direct impact of A on outcome Y, we take Y as the logarithmic weight gain of each pup at the end of study, compared with the baseline data. In addition, because the baseline weight is an important physical condition indicator of each pup and the principal stratum can be understood as the underlying healthy levels based on previous discussions, it is thus reasonable to consider A as a substitute for G in our motivating example.

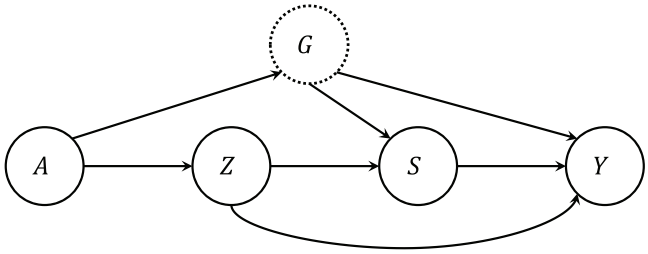

FIGURE 1 A causal diagram illustrating Assumptions 1 and 3. The solid circle represents observed variables, and the dotted circle represents latent principal stratum G. Note that observed covariates $\boldsymbol{C}$ can affect all variables in this plot and we omit it for simplicity

Theorem 1. *Suppose Assumptions 1–3 hold. If for any* $\boldsymbol{C} = \boldsymbol{c}$, *there exist* $L = m - z + 1$ *different values* $a_1, \ldots, a_L$ *of* A, *such that the* $L \times L$ *matrix* $\boldsymbol{M} = \{\pi_k(\boldsymbol{x}_l)\}_{kl}$ $(k = z, \ldots, m;\ l = 1, \ldots, L)$ *is of full rank for* $\boldsymbol{x}_l = (a_l, \boldsymbol{c}^{\mathrm{T}})^{\mathrm{T}}$, *then* $\mu_g(z)$ *is identifiable for* $g \geqslant z$.

The nonsingularity condition for matrix $\boldsymbol{M}$ in Theorem 1 requires the dependence between A and G, which is similar to the instrumental relevance assumption in instrumental variable analyses. In particular, this condition implies that there are at least L different levels of A related to the principal stratum G for each value of $\boldsymbol{C}$. Because $\pi_k(\boldsymbol{x})$ is identifiable, we can test the full rank condition by an estimate of $\pi_k(\boldsymbol{x})$. Specifically, if $\boldsymbol{X}$ has only low-dimensional discrete variables, then $\pi_k(\boldsymbol{x})$ can be estimated by their corresponding frequency ratios. Otherwise, we need to specify models for $\pi_k(\boldsymbol{x})$.

In the above context, we impose the additional Assumption 3 for identification. This assumption may be questionable if all covariates in $\boldsymbol{X}$ can directly affect the outcome. To relax Assumption 3 to some extent, we now consider a semiparametric linear model that includes all covariates. Specifically, we model Y in the group $Z = z$ by the following linear model allowing for interactions between Z and

G, and interactions between Z, G and $\boldsymbol{X}$:

$$E(Y \mid Z=z, G=k, \boldsymbol{X}=\boldsymbol{x}; \boldsymbol{\beta}_{zk}) = \beta_{zk,0} + \beta_{zk,1} a + \boldsymbol{\beta}_{zk,2}^{\mathrm{T}} \boldsymbol{c}, \quad (k = z, \ldots, m), \tag{5}$$

where $\boldsymbol{x} = (a, \boldsymbol{c}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\beta}_{zk} \equiv (\beta_{zk,0}, \beta_{zk,1}, \boldsymbol{\beta}_{zk,2}^{\mathrm{T}})^{\mathrm{T}}$ is a vector of unknown parameters. Note that Assumption 3 requires $\beta_{zk,1}$ to be zero in the above linear model.

Theorem 2. *Suppose that Assumptions 1–2 and model (5) hold. If $\{\pi_k(\boldsymbol{x}), \pi_k(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}}\}_{k=z}^{m}$ are linearly independent functions of $\boldsymbol{x}$, then $\mu_g(z)$ is identifiable for $g \geqslant z$.*

Because $\pi_k(\boldsymbol{x})$ is identifiable, we can assess the validity of the linear independence condition in Theorem 2 by observed data. This theorem implies that even when Assumption 3 is violated, we can still obtain the identifiability of $\mu_g(z)$, provided that the semiparametric linear model (5) holds. Verifications of model (5) or Assumption 3 based only on observed data are both challenging, because they involve a latent variable G. Nevertheless, since the linear model (5) is quite general for a continuous outcome, we suggest using this model in such cases.

3.2 | Estimation and inference

As shown earlier, $\mu_g(z)$ is nonparametrically identifiable based on an auxiliary variable or semiparametrically identifiable under a linear model setting. Although nonparametric estimation and inference can be obtained in the first case, it may not be applicable especially when the set of covariates is large due to the curse of dimensionality. In this section, we provide a unified estimation method for $\mu_g(z)$. To make progress, we posit a parametric model for principal stratum $G = k$ conditional on $\boldsymbol{X} = \boldsymbol{x}$:

$$\pi_k(\boldsymbol{x}) = \pi_k(\boldsymbol{x}; \theta), \qquad (k = 0, \ldots, m) \tag{6}$$

and a semiparametric model for outcome Y conditional on $Z = z$, $G = k$ and $\boldsymbol{X} = \boldsymbol{x}$:

$$\mu(z, k, \boldsymbol{x}) = \mu(z, k, \boldsymbol{x}; \boldsymbol{\beta}_{zk}), \qquad (k = z, \ldots, m), \tag{7}$$

where both θ and $\boldsymbol{\beta}_{zk}$ are vectors of unknown parameters. Define $\boldsymbol{\beta}_z = (\boldsymbol{\beta}_{zz}^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{zm}^{\mathrm{T}})^{\mathrm{T}}$, and let θ^0 and $\boldsymbol{\beta}_z^0$ denote the true parameters in (6) and (7), respectively.

The expression in (2) implies that for estimation of $\mu_g(z)$, we only need to estimate θ^0 and $\boldsymbol{\beta}_{zg}^0$. We first obtain an estimator of θ^0 by maximum likelihood estimation. Based on (3) and (6), we have the following equation:

$$\mathrm{pr}(S = 1 \mid Z = z, \boldsymbol{X} = \boldsymbol{x}) = \textstyle\sum_{k=z}^{m} \pi_k(\boldsymbol{x}; \theta).$$

The observed conditional log-likelihood function of $(Z, S, \boldsymbol{X})$ is thus given by

$$l(\theta; Z, S, \boldsymbol{X}) = \sum_{i=1}^{n} \sum_{z=1}^{m} \Big[I(Z_i = z, S_i = 1) \log \Big\{ \sum_{k=z}^{m} \pi_k(\boldsymbol{X}_i; \theta) \Big\} + I(Z_i = z, S_i = 0) \log \Big\{ \sum_{k=0}^{z-1} \pi_k(\boldsymbol{X}_i; \theta) \Big\} \Big],$$

where $I(\cdot)$ denotes an indicator function. Since estimation of θ via directly maximizing the observed likelihood function is difficult, we thus employ the EM algorithm to obtain a maximum likelihood estimator $\widehat{\theta}$ by specifying a multinomial logistic model for $\pi_k(\boldsymbol{x}; \theta)$. The computational details are given in the Supplementary Material.

Next we provide an estimating equation-based approach for estimation of $\boldsymbol{\beta}_z$ after obtaining $\widehat{\theta}$. Define

$$\omega_k(z, \boldsymbol{x}) = \mathrm{pr}(G = k \mid Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x}). \tag{8}$$

Under Assumptions 1 and 2, $\omega_k(z, \boldsymbol{x})$ is identifiable and $\omega_k(z, \boldsymbol{x}) = \pi_k(\boldsymbol{x}) / \{\sum_{k=z}^{m} \pi_k(\boldsymbol{x})\}$. In addition, the conditional expectation of Y given $(Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x})$ can be expressed as follows:

$$E(Y \mid Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x}) = \textstyle\sum_{k=z}^{m} \mu(z, k, \boldsymbol{x}) \omega_k(z, \boldsymbol{x}). \tag{9}$$

This, combined with (6) and (7), implies that $E\{H(Z, S, \boldsymbol{X}, Y; \theta^0, \boldsymbol{\beta}_z^0) \mid \boldsymbol{X}\} = 0$, where

$$H(Z, S, \boldsymbol{X}, Y; \theta, \boldsymbol{\beta}_z) = I(Z = z, S = 1) \Big\{ Y - \textstyle\sum_{k=z}^{m} \mu(z, k, \boldsymbol{X}; \boldsymbol{\beta}_{zk}) \omega_k(z, \boldsymbol{X}; \theta) \Big\} \tag{10}$$

and

$$\omega_k(z, \boldsymbol{x}; \theta) = \frac{\pi_k(\boldsymbol{x}; \theta)}{\sum_{k=z}^{m} \pi_k(\boldsymbol{x}; \theta)} \qquad (k = z, \ldots, m). \tag{11}$$

We then obtain a consistent estimator $\widehat{\boldsymbol{\beta}}_z$ by solving the estimating equations

$$\mathbb{P}_n \big\{ \boldsymbol{B}_z(\boldsymbol{X}; \widehat{\theta}) H(Z, S, \boldsymbol{X}, Y; \widehat{\theta}, \boldsymbol{\beta}_z) \big\} = \boldsymbol{0}, \tag{12}$$

where $\mathbb{P}_n(U) = \sum_{i=1}^{n} U_i / n$ for a generic variable U, and $\boldsymbol{B}_z(\boldsymbol{X}; \theta)$ is an arbitrary vector of functions of covariates $\boldsymbol{X}$ with dimension no smaller than that of $\boldsymbol{\beta}_z$. If the dimension of the user-specified function is larger than that of $\boldsymbol{\beta}_z$, we may adopt the generalized method of moments (GMM)

(Hansen, 1982) for estimation. Then in view of (2), we can obtain our proposed estimator $\widehat{\mu}_g(z)$ of $\mu_g(z)$ as follows:

$$\widehat{\mu}_g(z) = \frac{\mathbb{P}_n\{\pi_g(\boldsymbol{X};\widehat{\theta})\mu(z,g,\boldsymbol{X};\widehat{\boldsymbol{\beta}}_{zg})\}}{\mathbb{P}_n\{\pi_g(\boldsymbol{X};\widehat{\theta})\}}. \tag{13}$$

Let I_g denote the index set of $(\theta^{\mathrm{T}}, \boldsymbol{\beta}_{zg}^{\mathrm{T}})$ in $(\theta^{\mathrm{T}}, \boldsymbol{\beta}_z^{\mathrm{T}})$, and we obtain the following asymptotic normality of $\widehat{\mu}_g(z)$.

Theorem 3. *Suppose that the conditions in Theorem 1 or 2 hold and models (6)–(7) are correctly specified. Then under regularity conditions in the sense of Hansen (1982), $n^{1/2}\{\widehat{\mu}_g(z) - \mu_g(z)\}$ converges in distribution to $N(0, \sigma_{zg}^2)$ with $\sigma_{zg}^2 = \boldsymbol{\tau}_{zg}^{\mathrm{T}}\boldsymbol{\Omega}_{zg}\boldsymbol{\tau}_{zg}/\pi_g^2$, where*

$$\boldsymbol{\tau}_{zg} = \left[E\left\{\frac{\partial\pi_g(\boldsymbol{X};\theta^0)}{\partial\theta^{\mathrm{T}}}\mu(z,g,\boldsymbol{X};\boldsymbol{\beta}_{zg}^0)\right\},\right.$$

$$\left.E\left\{\pi_g(\boldsymbol{X};\theta^0)\frac{\partial\mu(z,g,\boldsymbol{X};\boldsymbol{\beta}_{zg}^0)}{\partial\boldsymbol{\beta}_{zg}^{\mathrm{T}}}\right\}\right]^{\mathrm{T}},$$

$\boldsymbol{\Omega}_{zg}$ is the $I_g \times I_g$ submatrix of $\boldsymbol{\Omega} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1^{\mathrm{T}}}$,

$$\boldsymbol{V} = \begin{bmatrix} E(\partial\boldsymbol{\psi}/\partial\theta^{\mathrm{T}}) & \boldsymbol{0} \\ E\{\partial(H\boldsymbol{B})/\partial\theta^{\mathrm{T}}\} & E\{\partial(H\boldsymbol{B})/\partial\boldsymbol{\beta}_z^{\mathrm{T}}\} \end{bmatrix},$$

$$\text{and} \quad \boldsymbol{\Sigma} = \begin{Bmatrix} E(\boldsymbol{\psi}\boldsymbol{\psi}^{\mathrm{T}}) & E(H\boldsymbol{\psi}\boldsymbol{B}^{\mathrm{T}}) \\ E(H\boldsymbol{B}\boldsymbol{\psi}^{\mathrm{T}}) & E(H^2\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}) \end{Bmatrix}$$

with the score function $\boldsymbol{\psi} = \partial l(\theta^0; Z, S, \boldsymbol{X})/\partial\theta$ at the true value θ^0, $H = H(Z, S, \boldsymbol{X}, Y; \theta^0, \boldsymbol{\beta}_z^0)$, and $\boldsymbol{B} = \boldsymbol{B}_z(\boldsymbol{X}; \theta^0)$.

Theorem 3 shows that the estimator obtained by (13) is asymptotically normal and $\sqrt{n}$-consistent. In order to construct confidence intervals for $\mu_g(z)$, we need to provide a consistent estimator for the asymptotic variance σ_{zg}^2. Such an estimator can be obtained by replacing the expectations in σ_{zg}^2 with their sample analogues. However, due to complexity issues in calculating derivatives, we suggest bootstrap procedures to obtain estimators of the variance.

3.3 | Bounds

When the identification assumptions are violated, the estimation of bounds of principal causal effects may be preferred (Cheng and Small, 2006; Grilli and Mealli, 2008; Imai, 2008; Lee, 2009). In this subsection, we derive bounds of $\mu_g(z)$ under Assumptions 1 and 2. According to (2), we only need to obtain lower and upper bounds of $\mu(z, g, \boldsymbol{x})$ for bounds of $\mu_g(z)$. We note from (9) that the average of Y conditional on $(Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x})$ is a weighted average of $\{\mu(z, k, \boldsymbol{x})\}_{k=z}^m$ with weights $\{\omega_k(z, \boldsymbol{x})\}_{k=z}^m$ defined in (8). Thus intuitively, $\mu(z, g, \boldsymbol{x})$ is bounded below by an average of the bottom $100 \times \omega_g(z, \boldsymbol{x})\%$ of values of Y conditional on $(Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x})$, and we denote this lower bound by $\mu^L(z, g, \boldsymbol{x})$. Similarly, the upper bound of $\mu(z, g, \boldsymbol{x})$, denoted by $\mu^U(z, g, \boldsymbol{x})$, is an average of the top $100 \times \omega_g(z, \boldsymbol{x})\%$ of values of Y conditional on $(Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x})$. To formally state these results, we let $Q_{z\boldsymbol{x}}(y)$ denote the distribution function of outcome Y conditional on $(Z = z, S = 1, \boldsymbol{X} = \boldsymbol{x})$, and define the q-quantile of $Q_{z\boldsymbol{x}}(y)$ as $y_{z\boldsymbol{x},q} = \inf\{y : Q_{z\boldsymbol{x}}(y) \geqslant q\}$. Let $L_{z\boldsymbol{x},q}(y)$ and $U_{z\boldsymbol{x},q}(y)$ denote the truncated distributions of $Q_{z\boldsymbol{x}}(y)$ at the lower q-quantile and upper q-quantile, respectively; that is,

$$L_{z\boldsymbol{x},q}(y) = \begin{cases} Q_{z\boldsymbol{x}}(y)/q & \text{if } y < y_{z\boldsymbol{x},q}, \\ 1 & \text{if } y \geqslant y_{z\boldsymbol{x},q}, \end{cases}$$

$$U_{z\boldsymbol{x},q}(y) = \begin{cases} 0 & \text{if } y < y_{z\boldsymbol{x},1-q}, \\ \{Q_{z\boldsymbol{x}}(y) - 1 + q\}/q & \text{if } y \geqslant y_{z\boldsymbol{x},1-q}. \end{cases}$$

Then the lower and upper bounds of $\mu(z, g, \boldsymbol{x})$ are, respectively, given by

$$\mu^L(z,g,\boldsymbol{x}) = \int y \,\mathrm{d}\, L_{z\boldsymbol{x},\omega_g(z,\boldsymbol{x})},$$

$$\text{and} \quad \mu^U(z,g,\boldsymbol{x}) = \int y \,\mathrm{d}\, U_{z\boldsymbol{x},\omega_g(z,\boldsymbol{x})}. \tag{14}$$

Based on (2) and (14), we immediately obtain the following result.

Theorem 4. *If Assumptions 1 and 2 hold, then $\mu_g^L(z)$ and $\mu_g^U(z)$ are sharp lower and upper bounds for $\mu_g(z)$, where*

$$\mu_g^L(z) = \pi_g^{-1}E\{\pi_g(\boldsymbol{X})\mu^L(z,g,\boldsymbol{X})\}$$

$$\text{and} \quad \mu_g^U(z) = \pi_g^{-1}E\{\pi_g(\boldsymbol{X})\mu^U(z,g,\boldsymbol{X})\}.$$

Let $\Delta_g^L(z_1, z_2) = \mu_g^L(z_1) - \mu_g^U(z_2)$ and $\Delta_g^U(z_1, z_2) = \mu_g^U(z_1) - \mu_g^L(z_2)$. Then $\Delta_g^L(z_1, z_2)$ and $\Delta_g^U(z_1, z_2)$ are the sharp lower and upper bounds of $\Delta_g(z_1, z_2)$, respectively.

Theorem 4 is not only applicable to observational studies when the ignorable treatment assignment assumption holds, but also serves as a complementary to the bound results derived by Wang et al. (2017) in the context of multiarm randomized trials. Specifically, Wang et al. (2017) derived sharp bounds of principal causal

effects without covariates for a binary outcome, whereas we provide bounds using covariates for a general outcome in Theorem 4. In addition, if these new results are applied to randomized trials, we can potentially obtain tighter bounds than those obtained from Wang et al. (2017). The following proposition provides necessary and sufficient conditions for how to obtain strictly tighter bounds when a binary covariate is utilized for adjustment. Let $\widetilde{\mu}_g^L(z)$ and $\widetilde{\mu}_g^U(z)$ denote the unadjusted lower and upper bound of $\mu_g(z)$, respectively. The specific expressions of these two bounds are provided in the Supplementary Material.

Proposition 1. *Under randomized and monotonicity assumptions, we have that* $\mu_g^L(z) \geqslant \widetilde{\mu}_g^L(z)$ *and* $\mu_g^U(z) \leqslant \widetilde{\mu}_g^U(z)$. *Let* Y *be a continuous or binary outcome, and we consider a scenario where a binary covariate* $\boldsymbol{X} \in \{0,1\}$ *is utilized for* $\mu_g^L(z)$ *and* $\mu_g^U(z)$. *If* $z < m$, *then*

(1) $\mu_g^L(z) > \widetilde{\mu}_g^L(z)$ *if and only if* $y_{z0,\omega_g(z,0)} \neq y_{z1,\omega_g(z,1)}$, *and when* Y *is binary, the if and only if condition further requires* $Q_{z\boldsymbol{x}}(0) \neq \omega_g(z,\boldsymbol{x})$ $(\boldsymbol{x} = 0,1)$, *where* $Q_{z\boldsymbol{x}}(y)$ *and* $y_{z\boldsymbol{x},q}$ *are defined earlier in this section and* $\omega_g(z,\boldsymbol{x})$ *is defined in (8);*

(2) $\mu_g^U(z) < \widetilde{\mu}_g^U(z)$ *if and only if* $y_{z0,1-\omega_g(z,0)} \neq y_{z1,1-\omega_g(z,1)}$, *and when* Y *is binary, the if and only if condition further requires* $1 - Q_{z\boldsymbol{x}}(0) \neq \omega_g(z,\boldsymbol{x})$ $(\boldsymbol{x} = 0,1)$.

Under randomized and monotonicity assumptions, $\mu_m(m) = E\{Y(m) \mid G = m\} = E(Y \mid Z = m, S = 1)$ is identifiable, and hence the adjusted and unadjusted bounds for $\mu_m(m)$ are equal. We exclude such a case by assuming $z < m$ in Proposition 1. Similar results to Proposition 1 are established in Long and Hudgens (2013) for the case where both the treatment Z and outcome Y are binary. This proposition provides practitioners a simple regime for bounds improvement in multiarm randomized trials. For example, to improve bounds of $\mu_g(z)$ in the case of a continuous outcome, one can simply choose a binary covariate $\boldsymbol{X}$ satisfying that $y_{z0,\omega_g(z,0)} \neq y_{z1,\omega_g(z,1)}$ or $y_{z0,1-\omega_g(z,0)} \neq y_{z1,1-\omega_g(z,1)}$. These conditions are testable, because $\omega_g(z,\boldsymbol{x})$ is identifiable and can be easily estimated under randomized and monotonicity assumptions. The verification details of conditions in Proposition 1 can be found in Section S3.3 of the Supplementary Material.

4 | SIMULATION STUDIES

In this section, we conduct simulation studies to evaluate the finite sample performance of the proposed estimators. We carry out the simulation in the following steps.

Step 1. We create a random sample $(Z, \boldsymbol{X}, S, Y)$ of size $n = 500$ (or 2000, 5000) as follows: we first generate Z according to $\text{pr}(Z = z) = 1/m$ for $z = 1, \ldots, m$, and generate covariates $\boldsymbol{X} = (1, A, \boldsymbol{C})^{\mathrm{T}}$, where $(A, \boldsymbol{C})^{\mathrm{T}}$ are joint normal with mean $(1,1)^{\mathrm{T}}$, unit variance, and correlation 0.5; under Assumption 2, we generate the principal stratum $G = k$ from a multinomial logistic model: $\pi_k(\boldsymbol{x};\theta) = \exp(\theta_k^{\mathrm{T}}\boldsymbol{x}) / \sum_{k=0}^{m} \exp(\theta_k^{\mathrm{T}}\boldsymbol{x})$ for $k = 0, \ldots, m$, where $\theta = (\theta_1^{\mathrm{T}}, \ldots, \theta_m^{\mathrm{T}})^{\mathrm{T}}$; then we let $S = 1$ if $G \geqslant Z$ and $S = 0$ otherwise; finally, we generate Y^* from a linear model: $Y^* \mid Z = z, G = k, \boldsymbol{X} = \boldsymbol{x} \sim N(\boldsymbol{\beta}_{zk}^{\mathrm{T}}\boldsymbol{x}, 0.5^2)$ for $z = 1, \ldots, m$, $k \geqslant z$, and let $Y = Y^*$ if $S = 1$.

Step 2. We apply the proposed approach in Sections 3.1 and 3.2 to the generated data set and estimate the principal causal effects $\Delta_g(z_1, z_2)$ within some principal stratum $G = g$.

Step 3. Repeat Steps 1 and 2 for 200 times.

Since there are four values of the treatment variable in our application, we consider $m = 4$ in the simulation study. The true values of parameters in Step 1 are set as follows: $\theta_0 = (0,0,0)^{\mathrm{T}}$, $\theta_1 = (2,1,2)^{\mathrm{T}}$, $\theta_2 = (2,1.2,1.5)^{\mathrm{T}}$, $\theta_3 = (2,1.4,2)^{\mathrm{T}}$ and $\theta_4 = (2,1.6,2.5)^{\mathrm{T}}$, and we consider two sets of true values of $\boldsymbol{\beta}_{zk}$: (I) $\boldsymbol{\beta}_{zk} = (-z + k, 0, 1)^{\mathrm{T}}$ so that Assumption 3 is satisfied; (II) $\boldsymbol{\beta}_{zk} = (-z + k, 1, 1)^{\mathrm{T}}$ under model (5).

For estimation in Step 2, we first obtain an estimate $\widehat{\theta}$ by maximum likelihood with the EM algorithm. Then we postulate linear models with $\boldsymbol{C}$ and $\boldsymbol{X}$ as predictors for the outcome in cases (I) and (II), respectively. We solve estimating equations in (12) with $\boldsymbol{B}_z(\boldsymbol{X};\widehat{\theta}) = \{\omega_z(z,\boldsymbol{X};\widehat{\theta}), \ldots, \omega_m(z,\boldsymbol{X};\widehat{\theta}), \boldsymbol{C}^{\mathrm{T}}\}^{\mathrm{T}}$ for case (I) and $\boldsymbol{B}_z(\boldsymbol{X};\widehat{\theta}) = \{\omega_z(z,\boldsymbol{X};\widehat{\theta}), \ldots, \omega_m(z,\boldsymbol{X};\widehat{\theta}), \boldsymbol{X}^{\mathrm{T}}\}^{\mathrm{T}}$ for case (II) to obtain estimators of $\{\boldsymbol{\beta}_{zk}\}_{k=z}^{m}$, where $\omega_k(z,\boldsymbol{x};\theta)$ is defined in (11) for $z = 1, \ldots, m$ and $k \geqslant z$. Finally, we obtain an estimator of $\Delta_g(z_1, z_2)$ according to (13) and (1), and employ bootstrap procedures with 200 replications to calculate its variance. In this simulation study, we choose causal estimands $\Delta_g(z_1, z_2)$ with $z_1 = 1$ and $z_2, g \in \{2,3,4\}$ for illustration purposes, and the performances of estimators for other estimands are similar.

We report the average bias, empirical standard error (SE), and 95% confidence interval coverage probabilities for estimators of $\Delta_g(1, z_2)$ based on 200 replications under sample sizes 500, 2000, and 5000. Table 2 shows simulation results for cases (I) and (II). The results in these two

TABLE 2 The average bias, empirical standard error (SE), and 95% coverage probability (CP) across 200 replications of the proposed estimators for $\Delta_g(z_1, z_2)$ in two cases at different sample sizes, respectively

	z_1	z_2	Sample size	Bias	SE	CP	Bias	SE	CP	Bias	SE	CP
				$g = 2$			$g = 3$			$g = 4$		
Case (I)			500	−0.41	2.86	0.79	−0.29	2.73	0.76	0.15	1.14	0.87
	1	2	2000	−0.15	1.50	0.88	−0.01	2.23	0.88	0.11	0.86	0.89
			5000	−0.03	1.07	0.94	0.01	1.13	0.95	0.01	0.35	0.94
			500				−0.40	2.52	0.79	0.12	1.04	0.89
	1	3	2000		NA		−0.01	2.02	0.85	0.13	0.83	0.90
			5000				0.01	0.97	0.94	−0.00	0.30	0.93
			500							−0.06	0.95	0.91
	1	4	2000		NA			NA		0.10	0.77	0.91
			5000							0.00	0.22	0.93
	z_1	z_2	Sample size	Bias	SE	CP	Bias	SE	CP	Bias	SE	CP
				$g = 2$			$g = 3$			$g = 4$		
Case (II)			500	−0.14	2.13	0.80	0.03	2.49	0.80	−0.07	1.06	0.86
	1	2	2000	−0.03	1.45	0.81	0.15	1.31	0.90	−0.05	0.73	0.93
			5000	0.01	0.97	0.91	0.09	0.82	0.91	−0.00	0.27	0.93
			500				−0.08	2.48	0.81	−0.15	0.96	0.91
	1	3	2000		NA		0.08	1.16	0.91	−0.06	0.70	0.93
			5000				−0.04	0.67	0.92	−0.01	0.23	0.93
			500							−0.18	0.88	0.88
	1	4	2000		NA			NA		−0.04	0.68	0.93
			5000							−0.03	0.18	0.95

Note: NA: not applicable for $\Delta_g(1, z_2)$ with $g < z_2$.

cases are similar. We first observe that for each fixed treatment level z_2 and sample size, the estimator of $\Delta_4(1, z_2)$ performs better than others. This is because the proportion of principal stratum $G = 4$ accounts for the largest and it can be well estimated away from zero. Using such a well-estimated quantity as the denominator, the proposed estimator in (13) for $\mu_4(z_2)$, and then for $\Delta_4(1, z_2)$, exhibits smaller bias and variance. Second, we observe that for each fixed principal stratum g and sample size, the performance of the estimator for $\Delta_g(1, z_2)$ improves as the treatment level z_2 increases. According to (13), the performance of our estimator mainly depends on how well $\boldsymbol{\beta}_{z_2 g}$ can be estimated. For estimation of $\boldsymbol{\beta}_{z_2 g}$, although the effective sample sizes (i.e., $\sum_{i=1}^{n} I(Z_i = z_2, S_i = 1)$) are approximately the same for all levels of z_2, there exist more mixed unknown components for smaller z_2 as indicated by (10). This would make the corresponding estimator more variable. Finally, we observe that as sample size increases, both biases and standard errors of the proposed estimators become smaller and the coverage probabilities are close to the nominal value. These results support the consistency and asymptotic results of estimators obtained by the proposed approach.

5 | APPLICATION TO NTP DATA

In this section, we apply the proposed approach to a real data set from the developmental toxicology experiment of antimony trioxide in rats and mice conducted by NTP (NTP, 2017). In this experiment, 800 Wistar Han rats and B6C3F1/N mice were randomly exposed to one of four different dose levels of antimony trioxide aerosol by whole body inhalation: 0, 3, 10, and 30 mg/m^3. The data of each pup include gender (male/female), species (rat/mouse), body weights every week for 2 years, and survival status. The data set does not include pretreatment weights for pups. However, because this is a chronic experiment, we assume that the first-week weights are not affected by their exposed doses, and hence, we treat their first-week weights as the baseline weights.

In our analysis, we consider the administered toxin levels as different treatment arms Z, which are coded as 1–4 corresponding to dose levels from low to high. For example, $Z = 1$ denotes the group with zero toxin dose level or the control group. Let $S = 1$, if a pup survived at the 2 years since the receipt of the treatment and $S = 0$ otherwise. We choose outcome Y to be the difference between a survived

TABLE 3 Estimate and 95% confidence interval of each π_k for NTP data set

Probability	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
π_k	0.28 (0.23, 0.33)	0.07 (0.00, 0.14)	0.11 (0.02, 0.20)	0.20 (0.10, 0.30)	0.34 (0.27, 0.41)

pup's logarithmic weight at the end of 2 years and its baseline logarithmic weight. Thus, the outcome Y is undefined if a pup died within 2 years. We are interested in evaluating causal effects of toxin levels on pups' weights reduction.

5.1 | Estimation under the monotonicity assumption

We first analyze the data set under the monotonicity assumption. In fact, a simple descriptive statistical analysis implies that as the toxin level increases, the survival rate of pups decreases. This does not contradict with monotonicity assumption 2. According to the expressions in (4), we estimate each principal stratum probability π_k ($k = 0, \dots, 4$) and its corresponding 95% confidence interval. The results are shown in Table 3. We see that 28% of pups will die and 34% of them will be alive, no matter what level of toxin is assigned; for the remaining 38% of pups, their survival status will be affected by the amount of toxin levels in 2 years. We focus on the average causal effect comparing each treatment level with the control within the always-survival group, that is, $\Delta_4(z_1, z_2)$ for $z_1 = 2, 3, 4$ and $z_2 = 1$.

We next consider three scenarios: (i) with Assumption 3; (ii) with model (5); and (iii) without Assumption 3 or model (5). In scenario (i), we choose A to be the baseline logarithmic weights and let $\boldsymbol{C}$ denote the other covariates. It is likely that the weight difference, that is, the outcome Y, is not directly affected by the baseline weights. Such a way of choosing A can be similarly found in Ding et al. (2011) and Wang et al. (2017). We calculated the partial correlation between the baseline weight A and the weight difference Y after adjusting for Z and $\boldsymbol{C}$, and the result is nearly null. We also made a plot between A and Y within each level of Z and $\boldsymbol{C}$, and we found that all correlations between A and Y in the plot are quite small (see Figure S1 in the Supplementary Material). These empirical evidences may also provide the validity of Assumption 3. We then postulate a linear model in $\boldsymbol{C}$ for the outcome in this scenario. In scenario (ii), we employ a linear model in both A and $\boldsymbol{C}$ for the outcome. Under these two different models, we use the unified approach presented in Section 3.2 for estimation. The approach involves EM-based maximum likelihood method and GMM estimation, both of which require optimization about unknown parameters. The convergence of these two procedures may depend on initial values. Therefore, we choose 10,000 initial values for optimization, obtain each estimator, and calculate its corresponding variance. We find that estimators with variances around the minimum one are close to each other, showing good convergence and stability, but estimators with large variances are unstable. We thus finally select the estimator with the minimum variance. Based on the estimator, we also calculate the 95% confidence interval of the always-survival average causal effect. Finally, when the identification conditions are removed as in scenario (iii), we provide estimation of bounds of the principal causal effects with and without covariate adjustment. All these results are shown in Table 4.

Table 4 shows that the proposed approach yields similar results in scenarios (i) and (ii), where the point estimates under these two scenarios exhibit the same sign and are relatively close for each principal stratum effect. Comparing them in terms of the estimated 95% confidence intervals, we find that the approach under the linear model with all covariates as predictors results in narrower confidence intervals. These results generally imply that if the always-survival pups are exposed to the highest level of toxins (i.e., 30 mg/m^3), then they will have lower weight than had they been given no toxins. In contrast, if these pups are exposed to more moderate toxin levels (i.e., 3 or 10 mg/m^3), then there is little evidence that these moderate toxins will impact the pups' weights. The bound results shown in Table 4 further support our findings. For example, since the bounds of $\Delta_4(2, 1)$ and $\Delta_4(3, 1)$ both include zero, the causal effects of lower toxin levels on pups' weights are not significant. We also observe that adjusting for a binary covariate (i.e., species) yields narrower bounds, compared with the bounds without covariates adjustment. We illustrate this by verifying conditions in Proposition 1. Note that the principal stratum effect $\Delta_4(z_1, z_2)$ is defined as $\mu_4(z_1) - \mu_4(z_2)$ for $z_1 = 2, 3, 4$ and $z_2 = 1$. We only need to focus on bounds of $\mu_4(z)$ for each level of z. In fact, after some calculations provided in Section S4.2 of the Supplementary Material, we find that $\widehat{y}_{z0,\omega_4(z,0)} \neq \widehat{y}_{z1,\omega_4(z,1)}$ for $z = 1, 2, 3$. According to Proposition 1, this implies that the adjusted lower bound $\mu_4^L(z)$ is larger than the unadjusted one $\widetilde{\mu}_4^L(z)$. Similarly, since $\widehat{y}_{z0,1-\omega_4(z,0)} \neq \widehat{y}_{z1,1-\omega_4(z,1)}$, the adjusted and unadjusted upper bounds satisfy $\mu_4^U(z) < \widetilde{\mu}_4^U(z)$ for $z = 1, 2, 3$. Finally, because $\mu_4(4)$ is identifiable, it can be pointly estimated. The bounds with and without adjustment for $\mu_4(4)$ should be equal. These results together indicate that adjusting for covariates can indeed sharpen the bound of $\Delta_4(z_1, z_2)$, as shown in Table 4. In

TABLE 4 The estimates of principal causal effects $\Delta_4(z_1, z_2)$ under various assumptions for NTP data set

Methods	$z_1 = 2$ versus $z_2 = 1$	$z_1 = 3$ versus $z_2 = 1$	$z_1 = 4$ versus $z_2 = 1$
	Estimates and 95% confidence intervals		
Estimation under Assumption 3	−0.04 (−0.24, 0.15)	−0.10 (−0.26, 0.06)	−0.36 (−0.51, −0.21)
Estimation under linear model	−0.05 (−0.19, 0.09)	−0.08 (−0.19, 0.03)	−0.29 (−0.40, −0.19)
	Lower and upper bounds		
Unadjusted bounds	(−0.46, 0.36)	(−0.50, 0.26)	(−0.45, −0.01)
Adjusted bounds	(−0.37, 0.23)	(−0.40, 0.15)	(−0.39, −0.11)

summary, lower levels of antimony trioxide (e.g., 3 mg/m^3) have little impact on the pups' weights, but a high level of antimony trioxide (e.g., 30 mg/m^3) can have a significant causal effect on pups' weights reduction, which may be useful information to laboratory scientists.

5.2 | Sensitivity analysis with violation of monotonicity assumption

In this subsection, we perform sensitivity analysis to assess the robustness of the empirical results when the monotonicity assumption is violated, as was usually done in truncation-by-death problems with a binary treatment. However, completely violating the monotonicity assumption in our setting with a four-level treatment is a lot more complicated than that in settings with a binary treatment. We thus consider a simpler version by allowing the monotonicity assumption violated only between any two adjacent treatment levels. This is reasonable in our example. If two toxicity levels are too close, the potential survival status of a pup under one dose level cannot determine the survival status under the other dose level. Thus, it is possible that a pup survived at some toxicity level, but it did not survive at a slightly lower toxicity level. If we additionally decreased the toxicity level, the pup may would survive again. Under the relaxed monotonicity assumption, three new principal strata, namely $defier_k$ for $k = 1, 2, 3$, were introduced in our empirical study (see Table S2 in the Supplementary Material). The principal stratum $G = defier_k$ is introduced as a consequence of monotonicity violations between treatment levels k and $k + 1$. Hence, the definition of $G = defier_k$ is similar to that of $G = k$ except that only values of $\{S(k), S(k + 1)\}$ are reversed. For example, $G = 1$ implies that $S(1) = 1, S(2) = 0$, but $G = defier_1$ implies that $S(1) = 0, S(2) = 1$; values of $S(3)$ and $S(4)$ in these two strata are both equal to 0. In this specific application, pups in the subgroup $G = 1$ survive only at the control level, whereas pups in $G = defier_1$ survive only at the toxicity level of 3 mg/m^3.

In parallel with the three additional principal strata, three sensitivity parameters are required for our analysis. Specifically, we introduce $\eta_k(\boldsymbol{x}) = \pi_{defier_k}(\boldsymbol{x})/\pi_k(\boldsymbol{x})$ to capture the deviation from monotonicity, where $\pi_{defier_k}(\boldsymbol{x}) = \text{pr}(G = defier_k \mid \boldsymbol{X} = \boldsymbol{x})$ for $k = 1, 2, 3$. Given these sensitivity parameters, one can show the identification of principal causal effects under a linear model. Details can be found in Section S5 of the Supplementary Material. For ease of presentation, we consider a simple case where these three parameters are equal to a constant, that is, $\eta_1(\boldsymbol{x}) = \eta_2(\boldsymbol{x}) = \eta_3(\boldsymbol{x}) = \eta$. The sensitivity parameter η can take value from 0 to ∞, and we have monotonicity if $\eta = 0$. Based on the discussions in Ding and Lu (2017), we may assume without loss of generality that $0 \leqslant \eta \leqslant 1$. Additionally, to ensure that proportions of the principal strata are all between 0 and 1, the range of η should be $[0, 0.283]$. Figure 2 presents how the estimated principal causal effects change as the sensitivity parameter η increases. We find that the estimated confidence intervals of the effect of moderate toxicity levels (i.e., 3 or 10 mg/m^3) on pups' weights almost always contain 0, and the estimated effects of the highest toxicity level (i.e., 30 mg/m^3) are always significantly negative. These results indicate that our previous findings are relatively robust to such a kind of deviation from monotonicity assumption. Scientific considerations may allow us to specify more accurate sensitivity parameters and thus obtain more precise conclusions about these effects.

6 | DISCUSSION

In this paper, we develop a principal stratification-based approach to address the truncation-by-death problem in multiarm studies with ordinal treatments. We establish the identifiability of principal causal effects via an auxiliary variable or under a linear model setting, and provide a unified semiparametric estimation approach. We also derive asymptotic normality results of our proposed estimator. Besides, when those identifiability conditions are not valid, we provide sharp bounds of the principal causal effects. In multiarm randomized trials, the derived bounds are narrower than those without covariates adjustment.

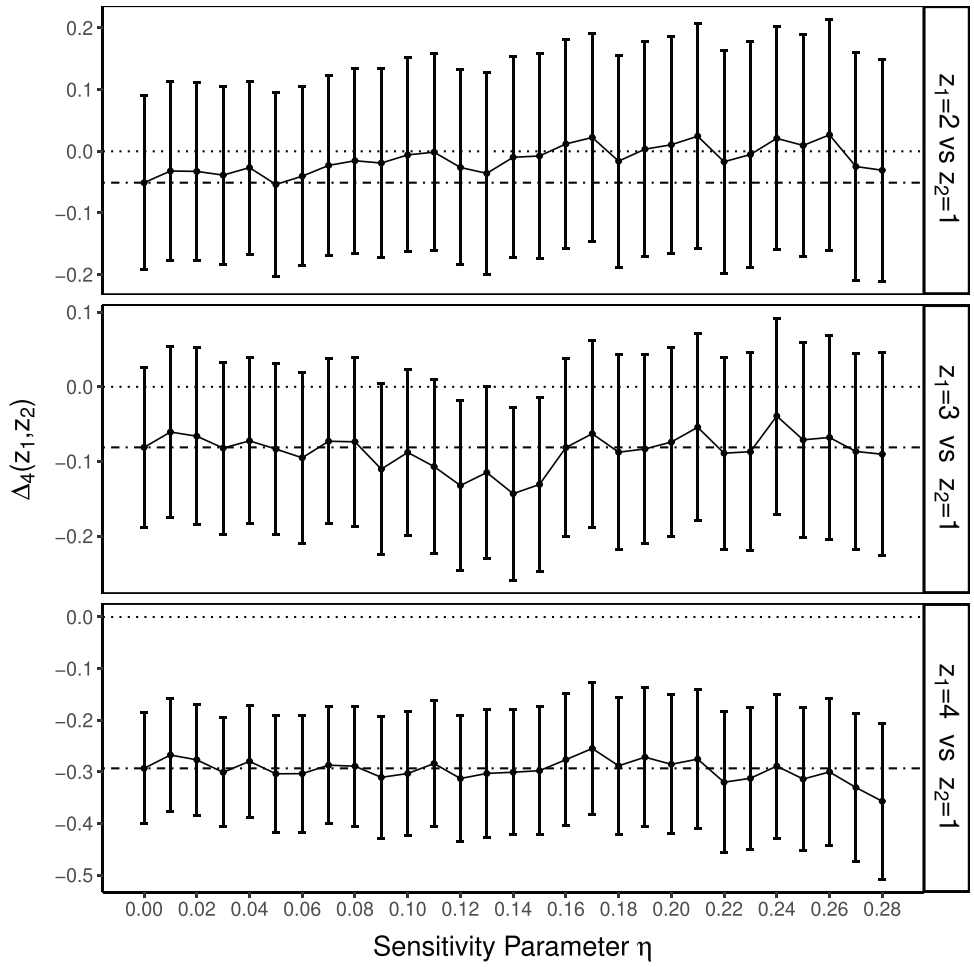

FIGURE 2 Sensitivity analysis for NTP data set with violation of monotonicity assumption. Point estimate and 95% confidence interval of the principal causal effect $\Delta_4(z_1, z_2)$ are shown for each value of the sensitivity parameter η. In each plot, the dot-dashed line indicates the point estimate under monotonicity assumption

The truncation-by-death problems have been extensively studied in the binary treatment setting (Ding et al., 2011; Tchetgen Tchetgen, 2014; Wang et al., 2017). Although we are still interested in pairwise comparisons of average potential outcomes in multiarm studies, one should not directly apply existing approaches to estimate the causal estimand by focusing analysis to the two treatment groups. In that manner, one can only obtain an estimate of a mixture of the causal parameter of interest defined in (1), because the principal stratum determined by potential survival status of two treatment levels is actually a coarse set of the basic principal stratum G. Therefore, it is necessary to develop approaches to address truncation-by-death problems in the multivalued treatment setting.

There are several possible directions for future research. First, the proposed approach may be generalized to deal with longitudinal and time-to-event data (Frangakis et al., 2004). Second, although we focus on a four-arm trial in our application, the proposed sensitivity analysis method is applicable to general multiarm studies. Third, since treatment variables of interest may be continuous in some truncation-by-death problems, it would be of interest to develop a principal stratification-based method to handle such settings. The study of these issues is beyond the scope of this paper and we leave them as future research topics.

ACKNOWLEDGMENTS

We sincerely thank the editor, associate editor, and reviewers for their helpful and insightful comments. Shanshan Luo and Yangbo He are supported by NSFC (11671020,11971040) and Huawei. Wei Li is supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China. The authors also express special thanks to Prof. Zhi Geng and Prof. Jinzhu Jia for their help and suggestions.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are openly available at https://cebs.niehs.nih.gov/cebs/publication/TR-590.

ORCID

Wei Li https://orcid.org/0000-0003-2461-2522

REFERENCES

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455.

Cheng, J. and Small, D.S. (2006) Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 815–836.

Chiba, Y. and VanderWeele, T.J. (2011) A simple method for principal strata effects when the outcome has been truncated due to death. *American Journal of Epidemiology*, 173, 745–751.

Ding, P., Geng, Z., Yan, W. and Zhou, X.-H. (2011) Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, 106, 1578–1591.

Ding, P. and Lu, J. (2017) Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 757–777.

Elliott, M.R., Joffe, M.M. and Chen, Z. (2006) A potential outcomes approach to developmental toxicity analyses. *Biometrics*, 62, 352–360.

Frangakis, C.E., Brookmeyer, R.S., Varadhan, R., Safaeian, M., Vlahov, D. and Strathdee, S.A. (2004) Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program. *Journal of the American Statistical Association*, 99, 239–249.

Frangakis, C.E. and Rubin, D.B. (2002) Principal stratification in causal inference. *Biometrics*, 58, 21–29.

Gilbert, P.B., Bosch, R.J. and Hudgens, M.G. (2003) Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*, 59, 531–541.

Grilli, L. and Mealli, F. (2008) Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. *Journal of Educational and Behavioral Statistics*, 33, 111–130.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

Hayden, D., Pauler, D.K. and Schoenfeld, D. (2005) An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*, 61, 305–310.

Imai, K. (2008) Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statistics & Probability Letters*, 78, 144–149.

Imbens, G.W. and Rubin, D.B. (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.

Lee, D.S. (2009) Training, wages, and sample selection: estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76, 1071–1102.

Lee, K., Daniels, M.J. and Sargent, D.J. (2010) Causal effects of treatments for informative missing data due to progression/death. *Journal of the American Statistical Association*, 105, 912–929.

Long, D.M. and Hudgens, M.G. (2013) Sharpening bounds on principal effects with covariates. *Biometrics*, 69, 812–819.

NTP (2017) TR-590: antimony trioxide (1309-64-4). Chemical effects in biological systems (CEBS). *Research Triangle Park, NC (USA): National Toxicology Program (NTP)*. Available from https://cebs.niehs.nih.gov/cebs/publication/TR-590.

Pearl, J. (2011) Principal stratification – a goal or a tool? *The International Journal of Biostatistics*, 7, 1–13.

Price, C.J., Kimmel, C.A., Tyl, R.W. and Marr, M.C. (1985) The developmental toxicity of ethylene glycol in rats and mice. *Toxicology and Applied Pharmacology*, 81, 113–127.

Rubin, D.B. (1990) Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5, 472–480.

Rubin, D.B. (2006) Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statistical Science*, 21, 299–309.

Shepherd, B.E., Gilbert, P.B., Jemiai, Y. and Rotnitzky, A. (2006) Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics*, 62, 332–342.

Tchetgen Tchetgen, E.J. (2014) Identification and estimation of survivor average causal effects. *Statistics in Medicine*, 33, 3601–3628.

Wang, L. and Richardson, T. (2017) On the concordant survivorship assumption. *Statistics in Medicine*, 36, 717–720.

Wang, L., Richardson, T.S. and Zhou, X.-H. (2017) Causal analysis of ordinal treatments and binary outcomes under truncation by death. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 719–735.

Wang, L., Zhou, X.-H. and Richardson, T.S. (2017) Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*, 104, 597–612.

Zhang, J.L., Rubin, D.B. and Mealli, F. (2009) Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104, 166–176.

SUPPORTING INFORMATION

Web Appendices, Figures, and Tables referenced in Sections 2-5 are available with this paper at the Biometrics website on Wiley Online Library. The R codes implementing the proposed approaches are also uploaded to the Biometrics repository.

Supporting Information.

How to cite this article: Luo, S., Li, W. and He, Y. (2021) Causal inference with outcomes truncated by death in multiarm studies. *Biometrics*, 1–12. https://doi.org/10.1111/biom.13554