Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Balancing Weights for Causal Inference in Observational Factorial Studies

Ruoqi Yu & Peng Ding

To cite this article: Ruoqi Yu & Peng Ding (08 Jan 2026): Balancing Weights for Causal Inference in Observational Factorial Studies, Journal of the American Statistical Association, DOI: 10.1080/01621459.2025.2602837

To link to this article: https://doi.org/10.1080/01621459.2025.2602837

Balancing Weights for Causal Inference in Observational Factorial Studies

Ruoqi Yu[a,*], Peng Ding[b]

[a]Department of Statistics, University of Illinois Urbana-Champaign

[b]Department of Statistics, University of California, Berkeley

*ruoqi.yu.ry@gmail.com

Abstract

Many scientific questions in biomedical, environmental, and psychological research involve understanding the effects of multiple factors on outcomes. While factorial experiments are ideal for this purpose, randomized controlled treatment assignment is generally infeasible in many empirical studies. Therefore, investigators must rely on observational data, where drawing reliable causal inferences for multiple factors remains challenging. As the number of treatment combinations grows exponentially with the number of factors, some treatment combinations can be rare or missing by chance in observed data, further complicating factorial effects estimation. To address these challenges, we propose a novel weighting method tailored to observational studies with multiple factors. Our approach uses weighted observational data to emulate a randomized factorial experiment, enabling simultaneous estimation of the effects of multiple factors and their interactions. Our investigations reveal a crucial nuance: achieving balance among covariates, as in single-factor scenarios, is necessary but insufficient for unbiasedly estimating factorial effects; balancing the factors is also essential in multi-factor settings. Moreover, we extend our weighting method to handle missing treatment combinations in observed data. Finally, we study the asymptotic behavior of the new weighting estimators and propose a consistent variance estimator, providing reliable inferences on factorial effects in observational studies.

Keywords: Covariate balance, factorial design, multiple factors, observational study, weighting.

1 Introduction

Assessing the effects of multiple factors is crucial in various scientific fields, such as biomedical research, environmental sciences, and psychology, as it helps shape decision-making and policy development. For instance, Berenson et al. (1998) used the Bogalusa Heart Study to investigate the effect of multiple cardiovascular risk factors on atherosclerosis in children and young adults, and Rillig et al. (2019) studied how ten global change factors affect soil properties, soil processes, and microbial biodiversity. A simple and common approach to evaluating the effects of multiple factors is to consider one factor at a time. However, this method requires considerable time and resources to estimate all effects of interest with the same precision, as investigators need to repeat the procedure multiple times. Moreover, evaluating one factor at a time limits our understanding of how the effect of one factor depends on other factors, which is critical in complex systems. For example, effective treatments of diseases such as sepsis, dementia, and stroke require combining treatments to target multiple components in cellular pathways (Berry et al., 2015).

To understand both the main effects and interactions of multiple factors, a powerful tool is the randomized factorial experiment involving random assignments of all possible treatment combinations to units (Wu and Hamada, 2021). Extensive research has been conducted on randomized factorial experiments (Box et al., 1978; Dasgupta et al., 2015; Dong, 2015; Branson et al., 2016; Lu, 2016; Mukerjee et al., 2018; Zhao et al., 2018; Egami and Imai, 2019; Kuhn et al., 2019; Li et al., 2020; Zhao and Ding, 2022; Pashley and Bind, 2023; Zhao and Ding, 2023). Although randomized experiments are considered the gold standard for estimating treatment effects, they are not always feasible due to ethical or practical considerations, forcing investigators to rely on observational data. However, the non-randomized treatment assignment mechanism in observational studies poses challenges for drawing reliable causal inferences, necessitating the removal of confounding due to observed covariates. Therefore, it is essential to explore how we can design observational studies (without examining any outcomes of interest) to emulate a randomized factorial experiment to estimate the effects of multiple factors and their interactions simultaneously, which is key to ensuring the objectivity of causal conclusions (Rubin, 2008).

Weighting is a popular approach for designing observational studies to emulate randomized experiments by removing confounding. Weighting aims to assign new weights to the individuals so that the weighted treatment groups are comparable in observed covariate distributions. Much of the literature on weighting has focused on one binary treatment. Widely used methods include inverse-probability weighting (Robins et al., 2000; Hirano and Imbens, 2001; Hirano et al., 2003) and balancing weighting (Hainmueller, 2012; Zubizarreta, 2015; Chan et al., 2016; Li et al., 2018; Wong and Chan, 2018; Zhao, 2019; Wang and Zubizarreta, 2020; Bruns-Smith and Feller, 2022; Cohn et al., 2023; Fan et al., 2023; Yu and Wang, 2024). A straightforward approach to generalize the weighting methods for one factor to multiple factors is to view the treatment combination as a single treatment with multiple levels, where researchers have made progress in recent years (McCaffrey et al., 2013; Fong and Imai, 2014; Li and Li, 2019; Resa and Zubizarreta, 2020; Shu et al., 2023). For a review of the relevant literature, see Lopez and Gutman (2017). By creating similar treatment combination groups, we can estimate all factorial effects of these factors, including all high-order interactions. This idea works well when the number of factors is relatively small and all treatment combinations are not rare. However, challenges arise with many factors or some

rare treatment combinations. Firstly, this approach can be computationally challenging due to the exponential growth of the number of treatment combinations as the number of factors increases. Moreover, when the number of factors is large, some treatment combinations are likely to be rare or empty, making this approach infeasible. When estimating all factorial effects is impossible, can we still draw inferences for the relatively important ones?

To answer this question, it is crucial to first determine which factorial effects are essential to estimate. Generally, low-order effects are considered more important than high-order effects, because the high-order effects are often believed to be smaller. This is known as the *effect hierarchy principle*. Additionally, factorial designs typically assume the *effect sparsity principle*, meaning that only a small number of relatively important effects are present (Box and Meyer, 1986). Thus, it is generally reasonable for investigators to assume that high-order interactions are negligible (i.e., essentially zero) and focus on estimating main effects and low-order interactions. For more discussions on these principles, see Wu and Hamada (2021). Although these effect principles originated from agricultural and industrial experiments, they are applicable to many other research fields. Still, investigators should validate their relevance using domain-specific knowledge when applying them to observational factorial studies. With these guiding principles, the next question is: how can we design an observational study to estimate the main effects and low-order interactions?

Towards this goal, we introduce a novel weighting approach designed explicitly for observational studies with multiple factors, as presented in §3. By employing a weighted sample to emulate a randomized factorial experiment, we can estimate the effects of multiple factors and their interactions simultaneously using the same set of weights without the need to design observational studies repeatedly for different factorial effects. Similar to the traditional balance weighting methods for a single binary treatment, balancing the observed covariates is necessary to obtain unbiased estimates of factorial effects. However, our investigations suggest that researchers must also balance the factors in the presence of multiple factors. We also study the asymptotic behavior of the new weighting estimators and propose a consistent variance estimator, which enables downstream statistical inference. The proposed weighting method easily generalizes to design incomplete observational factorial studies with incompletely observed treatment combinations, as discussed in §4. This extension is important when some treatment combination groups are empty, making it possible to deal with a large number of factors. Finally, we evaluate the performance of the proposed method via simulation in §5 and apply it to estimate the factorial effects of four volatile organic compounds on tachycardia using an observational study in §6.

2 Problem Setup and Weighting Framework

2.1 Problem Setting and Notation

Suppose we have $K \geq 2$ binary factors denoted by $z_k \in \{-1, +1\}$ for $k = 1, ..., K$. Then there are $Q = 2^K$ possible treatment combinations $\mathbf{z} = (z_1, \ldots, z_K)$. Let $\mathcal{Z} = \{-1, +1\}^K$ denote the set of possible treatment combinations. Our goal is to estimate the effects of multiple factors, including their interactions, under a full factorial design. To define the causal estimands, we follow the potential outcome framework for factorial designs proposed by Dasgupta et al. (2015). We denote the potential outcome if assigned to treatment combination $\mathbf{z} \in \mathcal{Z}$ as $Y(\mathbf{z})$. Each individual has *Q* potential outcomes; we denote the column vector of potential

outcomes as $\mathbf{Y} = \big(Y(\mathbf{z})\big)_{\mathbf{z}\in\mathcal{Z}}$. Also, let $\mathbb{E}[\mathbf{Y}] = \big(\mathbb{E}[Y(\mathbf{z})]\big)_{\mathbf{z}\in\mathcal{Z}}$ denote the corresponding column vector of expectations.

We define the causal estimands as in a randomized factorial experiment with a balanced design, where the treatment assignment is not related to any individual characteristics and each unit has an equal probability of receiving any treatment combination. Specifically, the main effect of each treatment z_k, τ_k, is defined as the comparison between the average potential outcomes of receiving treatment z_k and the average potential outcomes of not receiving it, uniformly averaged over all treatment combinations of other factors. More precisely, we can write τ_k as a contrast of the expected potential outcomes $\mathbb{E}[\mathbf{Y}]$:

$$\tau_k = \frac{1}{2^{K-1}} \mathbf{g}_k^{\mathrm{T}} \mathbb{E}[\mathbf{Y}],$$

where the contrast vector $\mathbf{g}_k = (g_{k\mathbf{z}})_{\mathbf{z}\in\mathcal{Z}}$ is a 2^K-dimensional vector with half $+1$s and half -1 s, indicating whether the treatment combination $\mathbf{z}$ has $z_k = +1$ or $z_k = -1$. Additionally, the interaction effect of two factors z_k and $z_{k'}$, denoted as $\tau_{k,k'}$, measures the difference between the average effect of receiving z_k versus not receiving it when $z_{k'}$ is received ($z_{k'} = +1$) or not ($z_{k'} = -1$), uniformly averaged over all treatment combinations of the other factors. Following Dasgupta et al. (2015), we define the corresponding contrast vector as $\mathbf{g}_{k,k'} = \mathbf{g}_k \circ \mathbf{g}_{k'}$, where $\circ$ denotes the component-wise product. Then, this interaction effect can be written as

$$\tau_{k,k'} = \frac{1}{2^{K-1}} \mathbf{g}_{k,k'}^{\mathrm{T}} \mathbb{E}[\mathbf{Y}].$$

We can define higher-order interactions in a similar way, for which the contrast vector equals to the component-wise product of the corresponding contrast vectors for the main effects. In general, with $[K] = \{1, \ldots, K\}$ denoting the set of indices for all factors, let $\tau_{\mathcal{K}}$ denote the factorial effect for treatments in $\mathcal{K} \subseteq [K]$ and let $\mathbf{g}_{\mathcal{K}} = (g_{\mathcal{K}\mathbf{z}})_{\mathbf{z}\in\mathcal{Z}}$ denote the corresponding contrast vector.

To estimate the factorial effects, suppose there are *N* subjects in the observational study, with $N_{\mathbf{z}}$ individuals with treatment combination $\mathbf{z}$. We assume $(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \ldots, N$, are independent and identically distributed, where $\mathbf{Z}_i \in \mathcal{Z}$ denotes the *K*-dimensional received factors, $\mathbf{X}_i = (X_{i1}, \ldots, X_{iD})$ denotes the *D*-dimensional covariate with density $f(\mathbf{X})$, and $\mathbf{Y}_i$ denotes the *Q*-dimensional column vector of potential outcomes for unit *i*. Additionally, we assume the *ignorability* of the treatment assignment conditional on the observed covariates $\mathbf{X}$: $0 < \mathbb{P}(\mathbf{Z}_i = \mathbf{z} | \mathbf{X}_i) < 1$ and $Y_i(\mathbf{z}) \perp\!\!\!\perp \mathbf{Z}_i | \mathbf{X}_i$ for all $\mathbf{z} \in \mathcal{Z}$ (Rosenbaum and Rubin, 1983). Since the covariate distribution may differ across treatment groups in an observational study, let $f_{\mathbf{z}}(\mathbf{X}) = f(\mathbf{X} | \mathbf{Z} = \mathbf{z})$ denote the covariate distribution in group $\mathbf{z}$. We can write the observed outcome for unit *i* as $Y_i^{\mathrm{obs}} = Y_i(\mathbf{Z}_i) = \sum_{\mathbf{z}\in\mathcal{Z}} I(\mathbf{Z}_i = \mathbf{z}) Y_i(\mathbf{z})$, where $I(\mathbf{Z}_i = \mathbf{z})$ is an indicator for whether unit *i* received treatment combination $\mathbf{z}$.

Although each individual has Q potential outcomes, we cannot observe the potential outcome $Y_i(\mathbf{z})$ if $\mathbf{Z}_i \neq \mathbf{z}$. Moreover, the treatment assignment is not randomized, and may depend on observed covariates $\mathbf{X}_i$. With these challenges, how can we estimate all factorial effects of interest?

2.2 Weighting for Observational Factorial Studies

To estimate the treatment effects of multiple factors and their interactions under a factorial structure, we use the weighting framework, which is a popular tool to deal with one binary factor ($K=1$ in our setting). As discussed in Section 2.1, each factorial effect $\tau_{\mathcal{K}}$ can be written as a contrast of all expected potential outcomes, with the corresponding contrast vector $\mathbf{g}_{\mathcal{K}} = (g_{\mathcal{K}\mathbf{z}})_{\mathbf{z}\in\mathcal{Z}}$. Following Han and Rubin (2021), we write $g_{\mathcal{K}\mathbf{z}}$, the contrast coefficient of the expected potential outcome under treatment combination $\mathbf{z}$, as $g_{\mathcal{K}\mathbf{z}} = g^{+}_{\mathcal{K}\mathbf{z}} - g^{-}_{\mathcal{K}\mathbf{z}}$, where $g^{+}_{\mathcal{K}\mathbf{z}} = \max(g_{\mathcal{K}\mathbf{z}}, 0)$ and $g^{-}_{\mathcal{K}\mathbf{z}} = \max(-g_{\mathcal{K}\mathbf{z}}, 0)$. We decompose the factorial effect $\tau_{\mathcal{K}}$ into two parts $\tau_{\mathcal{K}} = \tau^{+}_{\mathcal{K}} - \tau^{-}_{\mathcal{K}}$, where

$$\tau^{+}_{\mathcal{K}} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{+}_{\mathcal{K}\mathbf{z}} \mathbb{E}[Y(\mathbf{z})] \quad \text{and} \quad \tau^{-}_{\mathcal{K}} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{-}_{\mathcal{K}\mathbf{z}} \mathbb{E}[Y(\mathbf{z})]$$

correspond to the average expected potential outcomes of the positive and negative parts, respectively. Estimating $\tau^{+}_{\mathcal{K}}$ and $\tau^{-}_{\mathcal{K}}$ follows the same logic, so by symmetry, we mainly focus on estimating $\tau^{+}_{\mathcal{K}}$ in the following discussion.

Under ignorability, we have

$$\tau^{+}_{\mathcal{K}} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{+}_{\mathcal{K}\mathbf{z}} \int \mathbb{E}[Y(\mathbf{z}) |\ \mathbf{X}] f(\mathbf{X}) \mathrm{d}\mathbf{X} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{+}_{\mathcal{K}\mathbf{z}} \int \mathbb{E}[Y |\ \mathbf{Z}=\mathbf{z}, \mathbf{X}] f(\mathbf{X}) \mathrm{d}\mathbf{X}.$$

Estimating the target parameter $\tau^{+}_{\mathcal{K}}$ is challenging because we do not observe $Y(\mathbf{z})$ for individuals from the population covariate distribution $f(\mathbf{X})$. With $N_{\mathbf{z}}$ individuals with outcome $Y(\mathbf{z})$ and covariate distribution $f_{\mathbf{z}}(\mathbf{X}) = f(\mathbf{X} |\ \mathbf{Z}=\mathbf{z})$, we introduce an additional term – the weighting function $w_{\mathbf{z}}(\mathbf{X}) = f(\mathbf{X}) / f_{\mathbf{z}}(\mathbf{X})$ – so that

$$\tau^{+}_{\mathcal{K}} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{+}_{\mathcal{K}\mathbf{z}} \int \mathbb{E}[Y |\ \mathbf{Z}=\mathbf{z}, \mathbf{X}] w_{\mathbf{z}}(\mathbf{X}) f_{\mathbf{z}}(\mathbf{X}) \mathrm{d}\mathbf{X}.$$

Consider the weighting estimator

$$\hat{\tau}^{+}_{\mathcal{K}} = \frac{1}{2^{K-1}} \sum_{\mathbf{z}\in\mathcal{Z}} g^{+}_{\mathcal{K}\mathbf{z}} \left[\frac{1}{N_{\mathbf{z}}} \sum_{i:\mathbf{Z}_i=\mathbf{z}} Y^{\mathrm{obs}}_i w_{\mathbf{z}}(\mathbf{X}_i) \right]. \quad (1)$$

Given the weights, the weighted average $\hat{\tau}^{+}_{\mathcal{K}}$ is the sample analogue of $\tau^{+}_{\mathcal{K}}$.

From equation (1), we need to assign a weight of $w_{\mathbf{Z}_i}(\mathbf{X}_i)$ to the observed outcome of individual *i*, determined by the specific treatment combination they received. Let $w_i = N w_{\mathbf{Z}_i}(\mathbf{X}_i)/(2^{K-1}N_{\mathbf{Z}_i})$ denote the normalized weights, and let $A^+_{i\mathcal{K}} = \sum_{\mathbf{z}\in\mathcal{Z}} g^+_{\mathcal{K}\mathbf{z}} I(\mathbf{Z}_i = \mathbf{z})$ and $A^-_{i\mathcal{K}} = \sum_{\mathbf{z}\in\mathcal{Z}} g^-_{\mathcal{K}\mathbf{z}} I(\mathbf{Z}_i = \mathbf{z}) = 1 - A^+_{i\mathcal{K}}$ denote whether individual *i* belongs to the positive or negative part of contrast $\mathbf{g}_{\mathcal{K}}$, respectively. We estimate the factorial effect $\tau_{\mathcal{K}}$ with the weighting estimator

$$\hat{\tau}_{\mathcal{K}} = \hat{\tau}^+_{\mathcal{K}} - \hat{\tau}^-_{\mathcal{K}}, \quad \text{where}\, \hat{\tau}^{\Omega}_{\mathcal{K}} = \frac{1}{N}\sum_{i=1}^{N} w_i A^{\Omega}_{i\mathcal{K}} Y_i^{\text{obs}}, \text{for} \Omega = +, -. \quad (2)$$

How can we estimate the weights w_i so that the same weights can be used to estimate multiple factorial effects simultaneously?

2.3 A Naive Weighting Method for Estimating All Factorial Effects

One straightforward approach to estimating the weights is to transform the problem with multiple factors into a problem with one multi-valued treatment by directly focusing on the treatment combinations. In doing so, the goal is to estimate $\mathbb{E}[Y_i(\mathbf{z})]$ for all $\mathbf{z} \in \mathcal{Z}$.

A common choice of weights is based on the generalized propensity score $\pi_{\mathbf{z}}(\mathbf{X}) = \mathbb{P}(\mathbf{Z} = \mathbf{z} | \mathbf{X})$, where $\mathbf{z} \in \mathcal{Z}$ and $\sum_{\mathbf{z}\in\mathcal{Z}} \pi_{\mathbf{z}}(\mathbf{X}) = 1$ (Imbens, 2000; Imai and Van Dyk, 2004). Specifically, the weights are chosen as the inverse probability of receiving the treatment combination, i.e., $w_i \propto 1/\pi_{\mathbf{Z}_i}(\mathbf{X}_i)$. As the generalized propensity score is typically unknown, it needs to be estimated in practice, for example, using multinomial logistic regression. However, this estimation process can lead to significant biases in factorial effects estimation due to the misspecification of the generalized propensity score model (Kang and Schafer, 2007). In addition, a large number of factors and extreme estimated probabilities can result in unstable estimates of the treatment effects.

In recent years, balancing weighting methods have been proposed to improve the robustness of weighting methods (Fong and Imai, 2014; Li and Li, 2019; Resa and Zubizarreta, 2020; Ben-Michael et al., 2024; Shu et al., 2023). Rather than estimating the generalized propensity score $\pi_{\mathbf{z}}(\mathbf{X})$, these methods estimate weights w_i directly by comparing some basis functions of the covariates (Fong and Imai, 2014), $h_s(\mathbf{X}), s = 1, \ldots, S$, which is usually implemented as

$$\left| \frac{1}{N}\sum_{i=1}^{N} w_i I(\mathbf{Z}_i = \mathbf{z}) h_s(\mathbf{X}_i) - \frac{1}{N}\sum_{i=1}^{N} h_s(\mathbf{X}_i) \right| \leq \Delta_s, \text{for} \mathbf{z} \in \mathcal{Z} \text{and} s = 1, \ldots, S,$$

where the Δ_s's are the tolerances of covariate imbalance. Ideally, to remove all biases in factorial effects estimation, we would set the tolerance levels Δ_s to zero, requiring exact balance among treatment combination groups. However, it may result in an infeasible optimization problem when the sample size is not large enough. Specifically, treatment combination sizes must satisfy the condition $N_{\mathbf{z}} \geq S \text{for all} \mathbf{z} \in \mathcal{Z}$ to ensure the existence of

such weights. As the number of factors *K* increases, it becomes harder to satisfy this condition, necessitating a larger sample size *N* to ensure that all treatment combination groups have enough samples. In cases where exact balance is unachievable due to the limited sample size, an alternative is to allow approximated balance by setting $\Delta_s > 0$. In such scenarios, choosing the tolerances Δ_s optimally is crucial in reducing the biases in factorial effects estimation while challenging in making the optimization problem feasible.

This idea of viewing the treatment combinations as a multi-valued treatment implicitly assumes we want to estimate all factorial effects. When unbiased estimation of all factorial effects is impossible, we take a different perspective and consider the following relaxed research question: Can we still draw reliable inferences for the relatively important ones?

In practice, researchers are often primarily interested in the main effects and low-order interactions up to order K' and assume high-order interactions are negligible. For instance, a canonical choice is $K' = 2$. Let $[K]_k = \{J \subseteq [K] : 0 < |J| \leq k\}$ denote the set of all *k*th-order and lower-order treatment interaction index subsets. Therefore, we assume that the outcome depends on treatment combination $\mathbf{z}$ only through treatment interactions in $[K]_{K'}$.

Assumption 1 .

The interaction effects with orders greater than K' are negligible, where $K' \leq K$.

Given the challenges discussed previously, how can we estimate the factorial effects of interest $\tau_{\mathcal{K}} = \mathbf{g}_{\mathcal{K}}^{\mathrm{T}} \mathbb{E}[\mathbf{Y}] / 2^{K-1}$ for all $\mathcal{K} \in [K]_{K'}$?

3 A New Weighting Approach for Estimating Factorial Effects

In this section, we introduce a new weighting approach to estimating the treatment effects of multiple factors and their non-negligible interactions under a factorial structure. As described in Section 2.1, each factorial effect can be expressed as a contrast of all expected potential outcomes. To illustrate the main idea of our new weighting method, we first consider in Section 3.1 the estimation of one factorial effect $\tau_{\mathcal{K}}$ for treatments in $\mathcal{K} \subseteq [K]$, with the corresponding contrast vector $\mathbf{g}_{\mathcal{K}} = (g_{\mathcal{K}\mathbf{z}})_{\mathbf{z} \in \mathcal{Z}}$. We discuss the simultaneous estimation of multiple factorial effects in Section 3.2. Besides factorial effect estimation, we also introduce a variance estimator for the effect estimation in Section 3.3, so that we can conduct inferences for these factorial effects.

3.1 Weighting for Estimating a Single Factorial Effect

3.1.1 General Additive Outcome Model

To describe the new weighting method, we start with a general additive model for the expected potential outcomes, which is formally described in Assumption 2. Specifically, we assume the expected potential outcomes are additive in the confounders $\mathbf{X}$ and treatment interactions up to order $K' \leq K$.

Assumption 2 .

The conditional expectation of potential outcomes follows a general additive outcome model: $\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i] = \mu(\mathbf{X}_i) + \nu(\mathbf{z})$, *where* h_s *'s are prespecified basis functions,* $\mu(\mathbf{X}_i) = \sum_{s=1}^{S} \alpha_s h_s(\mathbf{X}_i)$ *for some unknown* α_s *'s and* $\nu(\mathbf{z}) = \sum_{J \in [K]_{K'}} \beta_J \prod_{j \in J} z_j$ *for some unknown* β_J *'s.*

With this general additive outcome model in Assumption 2, we can decompose the conditional bias of estimating $\tau_{\mathcal{K}}$ into two components, $\mathbb{E}[\hat{\tau}_{\mathcal{K}}^{\Omega} - \tau_{\mathcal{K}}^{\Omega}|\ \{\mathbf{X}_i\}_{i=1}^N, \{\mathbf{Z}_i\}_{i=1}^N]$ for $\Omega = +, -$, where

$$\mathbb{E}\left[\hat{\tau}_{\mathcal{K}}^{\Omega} - \tau_{\mathcal{K}}^{\Omega}|\ \{\mathbf{X}_i\}_{i=1}^N, \{\mathbf{Z}_i\}_{i=1}^N\right] = \sum_{s=1}^{S} \alpha_s \left(\frac{1}{N} \sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} h_s(\mathbf{X}_i) - \mathbb{E}[h_s(\mathbf{X}_i)] \right)$$
$$+ \sum_{J \in [K]_{K'}} \beta_J \left(\frac{1}{N} \sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} \prod_{j \in J} Z_{ij} - \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \prod_{j \in J} z_j \right). \quad (3)$$

For detailed derivations of (3), see §S3.1 in the Supplementary Materials. The above analysis of bias decomposition offers two key insights. First, to control the bias, we need to balance the basis functions of the covariates in the positive and negative parts of a contrast. Although we may not know $\mathbb{E}[h_s(\mathbf{X}_i)]$ in practice, it is natural to replace it with its sample analogue $\sum_{i=1}^{N} h_s(\mathbf{X}_i) / N$. This leads to the covariate balance constraints

$$\sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} h_s(\mathbf{X}_i) = \sum_{i=1}^{N} h_s(\mathbf{X}_i), \quad \text{for} \Omega = +, - \text{and} s = 1, \ldots, S.$$

This requirement is similar to that for traditional balancing weighting methods. However, it is not enough to obtain an unbiased estimator of $\tau_{\mathcal{K}}$ with multiple factors. In addition to the different covariate distributions, the treatment distributions may also differ for each contrast. Since the factors can affect the outcome through other channels besides the factorial effect under consideration, we also need to control for the confounding from other factors. This intuition aligns with our bias decomposition above and suggests that we need to include the factors themselves and the non-negligible interactions as additional "covariates" for balancing such that

$$\frac{1}{N} \sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} \prod_{j \in J} Z_{ij} = \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \prod_{j \in J} z_j, \quad \text{for} \Omega = +, - \text{and} J \in [K]_{K'},$$

to emulate the treatment assignment mechanism in a balanced factorial design. These additional balance constraints can be simplified as

$$\sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} \prod_{j\in J} Z_{ij} = 0, \text{for} \Omega = +, - \text{and} J \in [K]_{K'}, J \neq \mathcal{K}.$$

A similar observation was mentioned in Chattopadhyay and Zubizarreta (2022) in the context of regression with multi-valued treatments. Notably, when there is only one factor (i.e., $K = 1$), this set of balancing constraints on factors is equivalent to the balancing constraints for a constant covariate, and, therefore, does not need to be considered separately. This observation is consistent with the traditional balancing weighting methods for one binary factor. However, with $K \geq 2$, we have more balancing constraints to consider. In addition, the number of balancing constraints is an increasing function of the highest order of non-negligible interactions K'. The smaller K' is, the easier the problem becomes.

3.1.2 Outcome Model with Treatment Effect Heterogeneity

The general additive outcome model in Assumption 2 covers a broad class of functions by allowing a flexible form of basis functions. However, this model separates the effects of treatments and covariates and does not account for treatment effect heterogeneity, which is common in many applications. To address this limitation, we consider a more general class of outcome models with treatment effect heterogeneity in Assumption 3, by allowing the coefficient of the basis functions to depend on the treatment combination $\mathbf{z}$. This is a subtle but critical relaxation that can yield more robust and reliable causal conclusions.

Assumption 3 .

The conditional expectation of potential outcomes follows an outcome model with treatment effect heterogeneity: $\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i] = \sum_{J\in[K]_{K'}} \mu(\mathbf{X}_i; \mathbf{z}_J)$, *where* $\mu(\mathbf{X}_i; \mathbf{z}_J)$ *belongs to the span of prespecified basis functions* $h_s(\mathbf{X}_i)$, $s = 1, \ldots, S$, *i.e.,* $\mu(\mathbf{X}_i; \mathbf{z}_J) = \sum_{s=1}^{S} \alpha_{sJ} \left(\prod_{j\in J} z_j \right) h_s(\mathbf{X}_i)$. *This characterization leads to* $\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i] = \sum_{s=1}^{S} \alpha_{s\mathbf{z}} h_s(\mathbf{X}_i)$, *where* $\alpha_{s\mathbf{z}} = \sum_{J\in[K]_{K'}} \alpha_{sJ} \prod_{j\in J} z_j$ *are unknown coefficients that depend on* $\mathbf{z}$.

Equivalently, we can write the heterogeneous treatment effect outcome model in Assumption 3 using new basis functions $q_{sJ}(\mathbf{X}_i, \mathbf{z}) = h_s(\mathbf{X}_i) \prod_{j\in J} z_j$ such that

$$\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i] = \sum_{s=1}^{S} \sum_{J\in[K]_{K'}} \alpha_{sJ} q_{sJ}(\mathbf{X}_i, \mathbf{z}) = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{q}(\mathbf{X}_i, \mathbf{z}), \quad (4)$$

where $\boldsymbol{\alpha} = \{\alpha_{sJ}\}_{s=1,\ldots,S,\ J\in[K]_{K'}}$ and $\mathbf{q}(\mathbf{X}_i, \mathbf{z}) = \{q_{sJ}(\mathbf{X}_i, \mathbf{z})\}_{s=1,\ldots,S,\ J\in[K]_{K'}}$. Similar to working with the general additive outcome model in Assumption 2, we analyze the bias decomposition to obtain the required balance constraints. The decompositions

$$\mathbb{E}[\hat{\tau}_{\mathcal{K}}^{\Omega} - \tau_{\mathcal{K}}^{\Omega} \mid \{\mathbf{X}_i\}_{i=1}^N, \{\mathbf{Z}_i\}_{i=1}^N] = \boldsymbol{\alpha}^{\mathrm{T}} \left[\frac{1}{N} \sum_{i=1}^N w_i A_{i\mathcal{K}}^{\Omega} \mathbf{q}(\mathbf{X}_i, \mathbf{Z}_i) - \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \mathbb{E}[\mathbf{q}(\mathbf{X}_i, \mathbf{z})] \right], \text{for} \Omega = +, -,$$

suggest that to achieve unbiased estimates of the factorial effect $\tau_{\mathcal{K}}$ under the heterogeneous treatment effect outcome model, we need to balance the new basis functions $q_{sJ}(\mathbf{X}_i, \mathbf{Z}_i) = h_s(\mathbf{X}_i) \prod_{j \in J} Z_{ij}$ for all $s = 1, \ldots, S$, $J \in [K]_{K'}$. By replacing $\mathbb{E}[\mathbf{q}(\mathbf{X}_i, \mathbf{z})]$ with its sample analogue, we require

$$\sum_{i=1}^N w_i A_{i\mathcal{K}}^{\Omega} \mathbf{q}(\mathbf{X}_i, \mathbf{Z}_i) = \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \sum_{i=1}^N \mathbf{q}(\mathbf{X}_i, \mathbf{z}), \quad \text{for} \Omega = +, -.$$

As in §3.1.1, the number of constraints $2S\left[\binom{K}{1} + \binom{K}{2} \ldots + \binom{K}{K'}\right]$ grows as the highest interaction order K' increases. The key difference due to the flexible form of the outcome model is that we must now balance the interactions between the treatment combinations and covariates, which results in more balance constraints. To highlight the difference between the general additive outcome model (Assumption 2) and the heterogeneous treatment effect outcome model (Assumption 3), we examine a simple scenario in which only the main effects are non-zero (i.e., $K' = 1$) in §S3.2 of the Supplementary Materials.

3.2 Weighting for Estimating Multiple Factorial Effects Simultaneously

Suppose the investigator believes interactions with order greater than $K' \leq K$ are negligible (Assumption 1). The goal is to estimate $Q_+ = \binom{K}{1} + \binom{K}{2} + \cdots + \binom{K}{K'}$ non-negligible factorial effects simultaneously. Specifically, we seek to estimate the factorial effect $\tau_{\mathcal{K}}$ with contrast vector $\mathbf{g}_{\mathcal{K}}$ for all $\mathcal{K} \in [K]_{K'} = \{J \subseteq [K] : 0 < \mid J \mid \leq K'\}$. Applying the derivations from §3.1.2 to each of the Q_+ factorial effects of interest leads to the following balance constraints that are necessary to obtain unbiased factorial effect estimates under the outcome model with treatment effect heterogeneity: for all $\mathcal{K} \in [K]_{K'}$ and $\Omega = +, -$,

$$\sum_{i=1}^N w_i A_{i\mathcal{K}}^{\Omega} \mathbf{q}(\mathbf{X}_i, \mathbf{Z}_i) = \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \sum_{i=1}^N \mathbf{q}(\mathbf{X}_i, \mathbf{z}),$$

where $\mathbf{q}(\mathbf{X}_i, \mathbf{Z}_i) = \{q_{sJ}(\mathbf{X}_i, \mathbf{Z}_i)\}_{s=1,\ldots,S,\, J \in [K]_{K'}}$ and $q_{sJ}(\mathbf{X}_i, \mathbf{Z}_i) = h_s(\mathbf{X}_i) \prod_{j \in J} Z_{ij}$.

Any weights that satisfy the balance constraints described above can provide unbiased estimates of all non-negligible factorial effects. Unlike most existing balancing weighting methods, which often allow both positive and negative weights, we focus on non-negative weights since the ideal weighting function $w_{\mathbf{z}}(\mathbf{X})$ is the ratio of two density functions. Non-negative weights also avoid extrapolation and therefore make the results more interpretable.

If there are several choices of non-negative weights $w_i \geq 0$ that satisfy the above balance constraints, which weight should we use? One standard solution is to minimize a convex measure of the weights, $m(w_i)$, to select the optimal set of weights. Investigators can choose $m(\cdot)$ based on their needs and preferences. For instance, entropy balancing weights proposed by Hainmueller (2012) use $m(x) = x\log x$; calibration weights in Chan et al. (2016) use $m(x) = R(x,1)$ to avoid extreme weights, where $R(x, x_0)$ is a continuously differentiable, non-negative, and strictly convex function for any fixed $x_0 \in \mathbb{R}$; Resa and Zubizarreta (2020) use $m(x) = x^2$ to minimize the variance of the treatment effects estimation.

In summary, to obtain the optimal factorial weights for estimating multiple factorial effects $\tau_{\mathcal{K}}, \mathcal{K} \in [K]_{K'}$, we solve the following optimization problem:

$$\begin{aligned} &\min_{w_i \geq 0: i=1,\ldots,N} \quad \sum_{i=1}^{N} m(w_i) \\ &\text{subject to} \quad \sum_{i=1}^{N} w_i A_{i\mathcal{K}}^{\Omega} \mathbf{q}(\mathbf{X}_i, \mathbf{Z}_i) = \frac{1}{2^{K-1}} \sum_{\mathbf{z} \in \mathcal{Z}} g_{\mathcal{K}\mathbf{z}}^{\Omega} \sum_{i=1}^{N} \mathbf{q}(\mathbf{X}_i, \mathbf{z}), \text{for} \Omega = +, - \text{and} \mathcal{K} \in [K]_{K'}. \end{aligned} \quad (5)$$

In general, problem (5) is feasible with high probability. See Proposition S1 in the Supplementary Materials §S4.1 for a formal statement and proof.

To efficiently solve the constrained convex optimization problem (5), we consider its dual form as an unconstrained concave maximization problem. We first write the linear constraints of problem (5) as $\mathbf{Bw} = \mathbf{b}$, where the matrix $\mathbf{B}$ has *i*th column $\mathbf{B}_i$ representing all balance criteria evaluated at individual *i* and $\mathbf{b} = \sum_{i=1}^{N} \mathbf{b}_i$. This leads to the dual problem

$$\max_{\gamma \geq 0, \lambda} \mathcal{L}^*(\lambda, \gamma), \quad (6)$$

where $\mathcal{L}^*(\lambda, \gamma) = \sum_{i=1}^{N} \{\rho(\gamma_i - \lambda^{\mathrm{T}} \mathbf{B}_i) - \lambda^{\mathrm{T}} \mathbf{b}_i\}$, with $\rho(v) = m((m')^{-1}(v)) - v(m')^{-1}(v)$. For the derivations of the form of the dual problem, see §S3.3 in the Supplementary Materials. Since the dual problem (6) has no other constraints except for the non-negativity constraints of the Lagrangian multipliers γ, it is easier to solve compared with the original problem (5).

Suppose the dual problem (6) has optimal solution $(\lambda, \hat{\gamma})$. Then we can obtain the optimal factorial weights using the closed-form expression

$$\hat{w}_i = (m')^{-1}(\hat{\gamma}_i - \lambda^{\mathrm{T}} \mathbf{B}_i).$$

Details of this derivation are provided in §S3.3 in the Supplementary Materials.

To control the variance of the factorial effects estimates, we follow Zubizarreta (2015) and consider a special choice $m(w_i) = w_i^2$ in the objective function of the original problem (5),

which gives the optimal factorial weights $\hat{w}_i = (\hat{\gamma}_i - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i)/2$. When $m(x) = x^2$, the dual problem becomes

$$\max_{\boldsymbol{\gamma}\geq 0,\boldsymbol{\lambda}} \mathcal{L}^{*}(\boldsymbol{\lambda},\boldsymbol{\gamma}) = \sum_{i=1}^{N}\left\{-\frac{1}{4}(\gamma_i - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i)^2 - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}_i\right\}. \quad (7)$$

Finding the optimal solution $(\boldsymbol{\lambda},\hat{\boldsymbol{\gamma}})$ of the dual problem can be computationally expensive, since the number of decision variables in problem (7) increases linearly with the sample size *N*. To improve the computational efficiency, we can simplify the dual problem (7) further by exploiting the non-negativity conditions of $\hat{\boldsymbol{\gamma}}$. Specifically, the Karush-Kuhn-Tucker conditions allow us to express $\hat{\boldsymbol{\gamma}}$ explicitly as $\hat{\gamma}_i = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i I\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i \geq 0\right)$. As a result, we can reformulate the dual problem (7) as the following non-constrained optimization problem:

$$\max_{\boldsymbol{\lambda}} \mathcal{L}^{**}(\boldsymbol{\lambda}) = \sum_{i=1}^{N}\left\{-\frac{1}{4}(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i)^2 I(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0) - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}_i\right\}. \quad (8)$$

The simplified problem in (8) can be efficiently solved using numerical optimization methods such as Newton's method and gradient methods.

Weights obtained by solving original problem (5) or its dual problem (8) yield $\sqrt{N}$-consistent estimates for the factorial effects under mild regularity conditions. The formal results are presented in Theorem 1 below; see §S4.3 in the Supplementary Materials for the proof.

Theorem 1 .

Under Assumptions 1, 3 and S1, the weighting estimator $\hat{\tau}_{\mathcal{K}}$ *with weights* $\mathbf{w}$ *solving problem (5) with* $m(x) = x^2$ *is asymptotically normal for any* $\mathcal{K} \in [K]_{K'}$*, i.e.,*

$$\sqrt{N}\left(\hat{\tau}_{\mathcal{K}} - \tau_{\mathcal{K}}\right) \to \mathcal{N}\left(0, \sigma^{*2}_{\mathcal{K}}\right), \text{ with } \sigma^{*2}_{\mathcal{K}} = \sigma^2_{\mathcal{K}} + \sigma^2_{\epsilon},$$

where

$$\sigma^2_{\mathcal{K}} = \frac{1}{2^K}\mathrm{Var}\left(\sum_{\mathbf{z}\in\mathcal{Z}} g_{\mathcal{K}\mathbf{z}}\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i]\right) \quad \text{and} \quad \sigma^2_{\epsilon} = \frac{\bar{\sigma}^2}{4}\boldsymbol{\lambda}^{*\mathrm{T}}\mathbb{E}\left[\mathbf{B}_i\mathbf{B}_i^{\mathrm{T}} I(\boldsymbol{\lambda}^{*\mathrm{T}}\mathbf{B}_i < 0)\right]\boldsymbol{\lambda}^{*}$$

with $\boldsymbol{\lambda}^{*} = \mathrm{argmax}_{\boldsymbol{\lambda}}\mathbb{E}[-\frac{1}{4}(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i)^2 I(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0) - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}_i]$ *and* $\bar{\sigma}^2$ *defined in Assumption S1.*

Based on Theorem 1, our weighting estimator $\hat{\tau}_{\mathcal{K}}$ is $\sqrt{N}$-consistent for the factorial effect $\tau_{\mathcal{K}}$. Moreover, the variance $\sigma^{*2}_{\mathcal{K}}$ depends on two components – the variance $\sigma^2_{\mathcal{K}}$ from the projection of potential outcomes to the covariate space and the variance σ^2_{ϵ} from the residuals of the projection. Similar variance decompositions occur in semiparametric

efficiency theory (Robins et al., 1994; Hahn, 1998; Chan et al., 2016). Notably, the same residual variance σ_ϵ^2 is relevant for estimating all factorial effects, while the projection variance $\sigma_{\mathcal{K}}^2$ varies according to the contrast vector $\mathbf{g}_{\mathcal{K}}$.

In addition, the weighting estimators for all non-negligible factorial effects follow a multivariate normal distribution asymptotically. This result, presented in Corollary 1, supports simultaneous inference of multiple factorial effects. The proof is straightforward and is hence omitted.

Corollary 1 .

Under Assumptions 1, 3 and S1, the weighting estimators $\{\hat{\tau}_{\mathcal{K}}\}_{\mathcal{K}\in[K]_{K'}}$ *with weights* $\mathbf{w}$ *solving problem (5) with* $m(x)=x^2$ *satisfy*

$$\sqrt{N}\left(\{\hat{\tau}_{\mathcal{K}}\}_{\mathcal{K}\in[K]_{K'}}-\{\tau_{\mathcal{K}}\}_{\mathcal{K}\in[K]_{K'}}\right)\to\mathcal{N}\left(\mathbf{0},\Sigma^*\right), \text{ with } \Sigma^*=\Sigma+\Sigma_\epsilon,$$

where the (s,t)*th entry of* Σ *equals* $\Sigma_{st}=\frac{1}{2^K}\mathrm{Cov}\left(\sum_{\mathbf{z}\in\mathcal{Z}} g_{\mathcal{K}_s\mathbf{z}}\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i],\sum_{\mathbf{z}\in\mathcal{Z}} g_{\mathcal{K}_t\mathbf{z}}\mathbb{E}[Y_i(\mathbf{z})|\ \mathbf{X}_i]\right)$

and all entries of Σ_ϵ *equals* σ_ϵ^2.

While the theoretical results presented here employ a specific choice of $m(x)=x^2$, similar properties remain valid for other selections of $m(\cdot)$, provided that $m(\cdot)$ is strictly convex and continuously differentiable. For analogous prerequisites concerning the objective function, see Chan et al. (2016). Regardless of the chosen $m(\cdot)$ function that meets these criteria, we can establish asymptotic normality under certain regularity conditions. The primary distinction arises in the asymptotic variance, where the residual variance σ_ϵ^2 adapts according to the choice of $m(\cdot)$.

A common technique in the weighting literature to further improve the performance of a weighting method is to combine it with outcome regression adjustments (Robins et al., 1994; Rosenbaum, 2002). This type of augmented estimator is also applicable to our balancing weighting method in a factorial design. For more discussions, see §S3.4 in the Supplementary Materials.

Throughout this paper, we focus on the standard factorial effects and the idea applies to general contrasts, as defined in Zhao and Ding (2022).

3.3 Variance estimation

To estimate the asymptotic variance of $\sqrt{N}(\hat{\tau}_{\mathcal{K}}-\tau_{\mathcal{K}})$, we adopt a similar approach to that in Chan et al. (2016). Define a combined parameter $\boldsymbol{\theta}=(\boldsymbol{\lambda}^{\mathrm{T}},t)^{\mathrm{T}}$ and $\eta_i(\boldsymbol{\theta})=(\psi_i'(\boldsymbol{\lambda}),\ w_i(A_{i\mathcal{K}}^{+}-A_{i\mathcal{K}}^{-})Y_i^{\mathrm{obs}}-t)^{\mathrm{T}}$, where $\psi_i'(\boldsymbol{\lambda})=-\frac{1}{2}\mathbf{B}_i\mathbf{B}_i^{\mathrm{T}}\boldsymbol{\lambda}I(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i<0)-\mathbf{b}_i$ is the first

derivative of $\psi_i(\boldsymbol{\lambda}) = -\frac{1}{4}(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i)^2 I(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0) - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}_i$ with respect to $\boldsymbol{\lambda}$ and the weights $w_i = (\gamma_i - \lambda^{\mathrm{T}}\mathbf{B}_i)/2 = -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i I(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0)/2$ is also a function of $\boldsymbol{\lambda}$. Recall that $\boldsymbol{\lambda} = \mathrm{argmax}_{\lambda} \frac{1}{N}\sum_{i=1}^{N}\psi_i(\boldsymbol{\lambda})$ and $\hat{\tau}_{\mathcal{K}} = \frac{1}{N}\sum_{i=1}^{N}\hat{w}_i A_{i\mathcal{K}}^{+} Y_i^{\mathrm{obs}} - \frac{1}{N}\sum_{i=1}^{N}\hat{w}_i A_{i\mathcal{K}}^{-} Y_i^{\mathrm{obs}}$. Then $\boldsymbol{\theta} = (\boldsymbol{\lambda}^{\mathrm{T}}, \hat{\tau}_{\mathcal{K}})^{\mathrm{T}}$ is the *Z*-estimator that satisfies the following estimating equation $N^{-1}\sum_{i=1}^{N}\eta_i(\boldsymbol{\theta}) = 0$. In addition, $\boldsymbol{\theta}^* = (\boldsymbol{\lambda}^{*\mathrm{T}}, \tau_{\mathcal{K}})^{\mathrm{T}}$ satisfies $\mathbb{E}[\eta_i(\boldsymbol{\theta}^*)] = 0$, where $\boldsymbol{\lambda}^* = \mathrm{argmax}_{\lambda}\mathbb{E}[\psi_i(\boldsymbol{\lambda})]$, $w_i^* = -\boldsymbol{\lambda}^{*\mathrm{T}}\mathbf{B}_i I(\boldsymbol{\lambda}^{*\mathrm{T}}\mathbf{B}_i < 0)/2$, and $\tau_{\mathcal{K}} = \mathbb{E}[w_i^*(A_{i\mathcal{K}}^{+} - A_{i\mathcal{K}}^{-})Y_i^{\mathrm{obs}}] = \frac{1}{2^{K-1}}\sum_{\mathbf{z}\in\mathcal{Z}} g_{\mathcal{K}\mathbf{z}}\mathbb{E}[Y_i(\mathbf{z})]$.

Therefore, the asymptotic variance of $\sqrt{N}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$ can be written as

$$\mathbf{H}^{*-1}\mathbb{E}\left[\eta_i(\boldsymbol{\theta}^*)\eta_i(\boldsymbol{\theta}^*)^{\mathrm{T}}\right]\mathbf{H}^{*-1}, \text{ with } \mathbf{H}^* = \mathbb{E}\left[\eta_i'(\boldsymbol{\theta}^*)\right],$$

where $\eta_i'(\boldsymbol{\theta}^*)$ is the first derivative of $\eta_i(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. For details, see the proof of Theorem 2 in §S4.4 in the Supplementary Materials.

Since we are only concerned with the asymptotic variance of $\sqrt{N}(\hat{\tau}_{\mathcal{K}} - \tau_{\mathcal{K}})$, we focus on estimating the last element of $\mathbf{H}^{*-1}\mathbb{E}[\eta_i(\boldsymbol{\theta}^*)\eta_i(\boldsymbol{\theta}^*)^{\mathrm{T}}]\mathbf{H}^{*-1}$. Therefore, we only need to consider the last row of $\mathbf{H}^{*-1}$. Then, we have an estimator for $\sigma_{\mathcal{K}}^{*2}$:

$$\hat{\sigma}_{\mathcal{K}}^{*2} = L^{\mathrm{T}}\left[\frac{1}{N}\sum_{i=1}^{N}\eta_i(\boldsymbol{\theta})\eta_i(\boldsymbol{\theta})^{\mathrm{T}}\right]L = \frac{1}{N}\sum_{i=1}^{N}\left(\eta_i(\hat{\theta})^{\mathrm{T}}L\right)^2,$$

where

$$L = \left[\left(-\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}_i^{\mathrm{T}}(A_{i\mathcal{K}}^{+} - A_{i\mathcal{K}}^{-})Y_i^{\mathrm{obs}} I\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0\right)/2\right)\left(-\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}_i\mathbf{B}_i^{\mathrm{T}} I\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{B}_i < 0\right)/2\right)^{-1}, -1\right]^{\mathrm{T}}$$

is a column vector with the same dimension as $\eta_i(\theta)$.

Theorem 2 .

Under Assumptions 1, 3 and S1, we have $\hat{\sigma}_{\mathcal{K}}^{*2} \xrightarrow{\mathbb{P}} \sigma_{\mathcal{K}}^{*2}$.

Theorem 2 suggests that $\hat{\sigma}_{\mathcal{K}}^{*2}$ is a consistent estimator of the asymptotic variance of $\sqrt{N}(\hat{\tau}_{\mathcal{K}} - \tau_{\mathcal{K}})$. See §S4.4 in the Supplementary Materials for the proof. Notably, the same variance estimation technique can be applied to provide a consistent estimator for the covariance matrix of multiple factorial effects. This can be achieved by considering a revised combined parameter $\boldsymbol{\theta} = (\boldsymbol{\lambda}^{\mathrm{T}}, (t_{\mathcal{K}})_{[K]_{\mathcal{K}'}}^{\mathrm{T}})^{\mathrm{T}}$ and working with the *Z*-estimator for

$\eta_i(\boldsymbol{\theta}) = (\psi_i'(\boldsymbol{\lambda})^{\mathrm{T}}, (w_i(A_{i\mathcal{K}}^{+} - A_{i\mathcal{K}}^{-})Y_i^{\mathrm{obs}} - t_{\mathcal{K}})_{[K]_{K'}}^{\mathrm{T}})^{\mathrm{T}}$. The detailed discussion is omitted due to the limited space.

4 Observational Factorial Studies with Incomplete Treatment Combinations

In the previous discussions, we defined factorial effects as contrasts of all expected potential outcomes based on the ideal scenario of a full factorial design, where all possible combinations of factors are randomized. We then estimated the factorial effects by emulating a randomized full factorial experiment. However, in some practical scenarios, it may not be possible to emulate a full factorial design, especially when the number of factors *K* is large, or some treatment combinations are rare or missing in the observed data. As the number of treatment combinations grows exponentially with the number of factors, it is likely that some treatment combinations are not observed due to limitations in units or cost. In such cases, how can we estimate the factorial effects of interest?

In this section, we extend the proposed balancing weighting method to emulate an incomplete factorial design (Byar et al., 1993; Pashley and Bind, 2023), which can refer to any subset of a full factorial design. Suppose there are Q_{u} unobserved treatment combinations. Let $\mathcal{Z}_{\mathrm{u}}$ denote this set of Q_{u} unobserved treatment combinations, and let $\mathcal{Z}_{\mathrm{o}}$ denote the set of $Q_{\mathrm{o}} = Q - Q_{\mathrm{u}}$ observed treatment combinations. When the interaction effects with order greater than K' are negligible (Assumption 1), we only need to estimate $Q_{+} = \binom{K}{1} + \binom{K}{2} + \cdots + \binom{K}{K'}$ factorial effects; all the other $Q_{-} = Q - Q_{+} - 1$ factorial effects are zero. How can we use the observed Q_{o} treatment combinations in $\mathcal{Z}_{\mathrm{o}}$ to infer the Q_{+} factorial effects?

To answer this question, we begin by exploring a fundamental question: Is it possible to identify the Q_{+} factorial effects using only Q_{o} observed treatment combinations? Let $\mathbf{G} = [\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_K, \mathbf{g}_{1,2}, \ldots, \mathbf{g}_{K-1,K}, \ldots, \mathbf{g}_{1,\ldots,K}]$ denote the design matrix for the average outcome across treatments and all $Q-1$ factorial effects, where the first column $\mathbf{g}_0 = (+1, \ldots, +1)^{\mathrm{T}}$. The corresponding vector of average outcome and factorial effects equals $\boldsymbol{\tau} = (2\tau_0, \tau_1, ..., \tau_K, \tau_{1,2}, ..., \tau_{1,...,K})^{\mathrm{T}} = \frac{1}{2^{K-1}}\mathbf{G}^{\mathrm{T}}\mathbb{E}[\mathbf{Y}]$. Under Assumption 1, the high-order interactions are negligible, so the last Q_{-} entries of $\boldsymbol{\tau}$ are zero. For convenience, we rearrange the entries of $\mathbb{E}(\mathbf{Y})$ and $\mathbf{G}$ such that the unobserved treatment combinations occur after the observed ones and the negligible contrasts occur after the non-negligible ones. Using tilde to denote a rearranged vector and matrix, we partition them as

$$\mathbb{E}(\tilde{\mathbf{Y}}) = \begin{bmatrix} (\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z}\in\mathcal{Z}_{\mathrm{o}}} \\ (\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z}\in\mathcal{Z}_{\mathrm{u}}} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_{\mathrm{o}+} & \mathbf{G}_{\mathrm{o}-} \\ \mathbf{G}_{\mathrm{u}+} & \mathbf{G}_{\mathrm{u}-} \end{bmatrix},$$

where $\mathbf{G}_{\mathrm{o}+}$ is the $Q_{\mathrm{o}} \times (Q_{+}+1)$ submatrix for the observed treatment combinations and non-negligible contrasts, $\mathbf{G}_{\mathrm{u}+}$ is the $Q_{\mathrm{u}} \times (Q_{+}+1)$ submatrix for the unobserved treatment combinations and non-negligible contrasts, $\mathbf{G}_{\mathrm{o}-}$ is the $Q_{\mathrm{o}} \times Q_{-}$ submatrix for the observed treatment combinations and negligible contrasts, $\mathbf{G}_{\mathrm{u}-}$ is the $Q_{\mathrm{u}} \times Q_{-}$ submatrix for the unobserved treatment combinations and negligible contrasts. Then $\boldsymbol{\tau} = \frac{1}{2^{K-1}} \tilde{\mathbf{G}}^{\mathrm{T}} \mathbb{E}(\tilde{\mathbf{Y}})$. Due to the orthogonality of $\tilde{\mathbf{G}}$ that $\tilde{\mathbf{G}}^{\mathrm{T}} \tilde{\mathbf{G}} = 2^{K} \mathbf{I}_{2^{K}}$, we can express the expected potential outcomes in terms of the factorial effects using the relationship $\mathbb{E}[\tilde{\mathbf{Y}}] = 2^{-1} \tilde{\mathbf{G}} \boldsymbol{\tau}$. This one-to-one correspondence between the factorial effects $\boldsymbol{\tau}$ and the expected potential outcomes $\mathbb{E}[\tilde{\mathbf{Y}}]$ allows us to identify all factorial effects in a full factorial design when all treatment combinations are observed. However, when some treatment combinations are unobserved, it is impossible to estimate the corresponding expected potential outcome without further assumptions. Consequently, we cannot estimate the factorial effects using the above one-to-one correspondence directly. Still, this relationship between $\boldsymbol{\tau}$ and $\mathbb{E}[\tilde{\mathbf{Y}}]$ is useful, as we can use it to infer the expected potential outcomes for the unobserved treatment combinations $(\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{u}}}$ by utilizing the negligibility of the Q_{-} highest-order interactions, given by

$$\mathbf{0} = \mathbf{G}_{\mathrm{o}-}^{\mathrm{T}} (\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{o}}} + \mathbf{G}_{\mathrm{u}-}^{\mathrm{T}} (\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{u}}}.$$

Specifically, if $\mathbf{G}_{\mathrm{u}-}$ has full row rank, we can infer the expected potential outcomes for unobserved treatment combinations $(\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{u}}}$ using the relationship

$$(\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{u}}} = -\mathbf{G}_{\mathrm{u}-} (\mathbf{G}_{\mathrm{u}-}^{\mathrm{T}} \mathbf{G}_{\mathrm{u}-})^{-1} \mathbf{G}_{\mathrm{o}-}^{\mathrm{T}} (\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z} \in \mathcal{Z}_{\mathrm{o}}}.$$

It is worth noting that having a full row rank $\mathbf{G}_{\mathrm{u}-}$ is a necessary and sufficient condition for identifying the non-negligible factorial effects. This condition is only possible to hold if the number of non-negligible factorial effects Q_{+} is no more than the number of observed treatment combinations Q_{o}.

Let $\boldsymbol{\tau}_{+}$ denote the vector of average outcome and non-negligible factorial effects. With the identified expected potential outcomes for the unobserved treatment combinations, we can identify $\boldsymbol{\tau}_{+}$ under an incomplete factorial design as in Proposition 1 below. Proposition 1 suggests that the factorial effects, originally defined based on the mean of all potential outcomes (the treatment combination groups contribute equally to each contrast), are equivalent to linear combinations of the expected potential outcomes for the observed treatment combinations. Therefore, different treatment combination groups can contribute to the factorial effects with different weights. To illustrate the idea, we consider a toy example with $K = 3$ binary treatments in §S3.5 in the Supplementary Materials. Notably, we can also relax the ignorability assumption in the sense that (i) the probabilistic assumption only needs to hold for the observed treatment combinations, and (ii) the unconfounded assumption only requires $\mathbf{Z}_i \perp\!\!\!\perp (Y_i(\mathbf{z}))_{\mathbf{z} \in \mathcal{Z}_{\mathrm{o}}} \mid \mathbf{X}_i$.

Proposition 1 .

If the matrix $\mathbf{G}_{\mathrm{u}-}$ *has full row rank, we have*

$$\boldsymbol{\tau}_{+}=\frac{1}{2^{K-1}}\mathbf{G}_{\mathrm{o}}^{\mathrm{T}}(\mathbb{E}[Y(\mathbf{z})])_{\mathbf{z}\in\mathcal{Z}_{\mathrm{o}}},\quad \text{where}\mathbf{G}_{\mathrm{o}}^{\mathrm{T}}=\mathbf{G}_{\mathrm{o}+}^{\mathrm{T}}-\mathbf{G}_{\mathrm{u}+}^{\mathrm{T}}\mathbf{G}_{\mathrm{u}-}(\mathbf{G}_{\mathrm{u}-}^{\mathrm{T}}\mathbf{G}_{\mathrm{u}-})^{-1}\mathbf{G}_{\mathrm{o}-}^{\mathrm{T}}.$$

With the formula for the non-negligible factorial effects in Proposition 1, we can formulate our optimization problem in a similar way to Section 3.2 to find the optimal weights that mimic an incomplete factorial design and hence can be used to estimate the non-negligible factorial effects. Although the problem formulation is similar as before, it is worth noting that the indicators of whether unit *i* contribute to the positive or negative parts of factorial effect $\tau_{\mathcal{K}}$ in a full factorial design, $A_{i\mathcal{K}}^{+}$ and $A_{i\mathcal{K}}^{-}$, have new meanings in an incomplete factorial design, representing the weights that unit *i* contributes to the positive or negative parts of factorial effect $\tau_{\mathcal{K}}$. Importantly, the theoretical properties and the form of the variance estimator remain unchanged, and we omit the proofs since they are the same as before.

5 Simulation

5.1 Performance of the Weighting Estimator

In the first set of simulations, we evaluate the performance of the proposed weighting estimator for estimating factorial effects. Suppose we observe three factors $\mathbf{Z}_i=(Z_{i1},Z_{i2},Z_{i3})$ and five covariates $\mathbf{X}_i=(X_{i1},\ldots,X_{i5})$ for each individual *i*. The covariates $\mathbf{X}_i$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu},\Sigma)$, where $\boldsymbol{\mu}=(0,0,0,0,0)^{\mathrm{T}}$ and Σ has diagonal elements 1 and off-diagonal elements ρ. The treatment assignment mechanism for Z_{ik} is independent across *k*'s and satisfies a logistic regression that $\mathbb{P}(Z_{ik}=1)=1/(1+\exp(-\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{X}_i))$, where $\boldsymbol{\beta}_1=(1/4,2/4,0,3/4,1),\boldsymbol{\beta}_2=(3/4,1/4,1,0,2/4),\boldsymbol{\beta}_3=(1,0,3/4,2/4,1/4)$. Here, we assume the conditional independence of factors given covariates for convenience, but it is not needed in the theory throughout the paper. This treatment assignment mechanism ensures all $2^3=8$ treatment combination groups are non-empty and are observed so that the proposed weighting estimators in Section 3 are applicable. Suppose only the main effects of the three treatments are non-negligible. We consider three outcome models: an additive outcome $Y_{i1}=2\sum_{j=1}^{5}X_{ij}+\sum_{k=1}^{3}Z_{ik}+\epsilon_{i1}$, a heterogeneous treatment effect outcome $Y_{i2}=2\sum_{j=1}^{5}X_{ij}+\sum_{k=1}^{3}Z_{ik}+\sum_{j=1}^{5}X_{ij}\sum_{k=1}^{3}Z_{ik}+\epsilon_{i2}$, and a misspecified outcome $Y_{i3}=4\sin(X_{i1})+\exp(0.4X_{i2}^2)+\left(\min(1,X_{i1})+X_{i2}\right)Z_{i1}+X_{i1}Z_{i2}+\sum_{j=1}^{5}X_{ij}Z_{i3}+\epsilon_{i3}$, where ϵ_{ij}'s are independent errors following a standard normal distribution. The true main effects for Y_1 are $\tau_1=2,\tau_2=2,\tau_3=2$. The true main effects for Y_2 are $\tau_1=2,\tau_2=2,\tau_3=2$. The true main effects for Y_3 are $\tau_1=2\mathbb{E}[\min(X_1,1)],\tau_2=0,\tau_3=0$. We consider four estimators for each main effect $\tau_k,k=1,2,3$, under each outcome model: (i) the additive regression estimator,

which is twice the coefficient of Z_k when regressing Y on the intercept, centralized X_j (i.e., $X_j - \bar{X}_j$) and Z_k for $j = 1, \ldots, 5, k = 1, 2, 3$, (ii) the interaction regression estimator, which is twice the coefficient of Z_k when regressing Y on the intercept, $X_j - \bar{X}_j$, Z_k and $(X_j - \bar{X}_j)Z_k$ for $j = 1, \ldots, 5, k = 1, 2, 3$ (Zhao and Ding, 2023), (iii) the proposed weighting estimator using balance constraints under the general additive model assumption (additive balance constraints) and covariate basis functions $h_s(\mathbf{X}) = X_s, s = 1, \ldots, 5$, and (iv) the proposed weighting estimator using balance constraints under the outcome model assumption with treatment effect heterogeneity (interaction balance constraints) and the same set of basis functions as (iii). We vary the sample size as $N = 500, 1000, 2000$ and vary the covariance as $\rho = 0.2, 0.4, 0.6$ for each scenario. In each simulation setting, we compare the absolute bias and root mean squared error (RMSE) using 1000 repetitions. We summarize the results for $N = 1000$ and $\rho = 0.4$ in Figure 1. We present the results for other choices of N and ρ and evaluate the effects of sample size N and covariance ρ in the Supplementary Materials §S1.

Figure 1 suggests that when the outcome is additive, the four estimators have similar performances and the regression estimators achieve slightly smaller RMSEs than the two weighting estimators. When the treatment effect is heterogeneous, the two methods allowing treatment effect heterogeneity (interaction regression estimator and weighting with interaction balance constraints) have smaller biases and RMSEs than the other two methods. When the outcome is misspecified, the weighting estimator with interaction balance constraints outperforms the other methods, achieving the smallest RMSEs. This observation confirms the robustness of the proposed weighting method.

In the Supplementary Materials §S1, we conduct additional simulations with a larger number of factors and under heteroskedasticity. We also consider another set of treatment assignment mechanism where not all treatment combination groups are observed, hence the proposed method in §3 is no longer applicable and the performance of the generalized method in §4 is evaluated.

5.2 Performance of the Variance Estimator

In the next set of simulations, we evaluate the performance of the proposed variance estimator in Theorem 2. We consider the same setting as the first set of simulations, with a sample size of $N = 500, 1000, 2000$ and $\rho = 0.4$. We estimate $\sqrt{N}(\hat{\tau}_k - \tau_k)$ in two ways: we obtain the benchmark estimator by adjusting the variance of estimated factorial effects by the sample size and also consider the average proposed consistent variance estimator across 1000 repetitions. From the results in Figure 2, we can see that the simulated variance estimator and the consistent variance estimator are similar, with a ratio getting closer to 1 as the sample size increases. Additionally, we calculate the empirical coverage probability of 95% confidence intervals obtained using normal distribution (Theorem 1) and the consistent variance estimator (Theorem 2). Results in Figure 2 suggest that the confidence intervals have coverage probabilities close to 0.95.

6 Application: VOCs Exposure and Tachycardia

Exposure to volatile organic compounds (VOCs) is pervasive in our daily life due to common indoor and outdoor sources such as building materials, home cleaning products, gasoline, and industrial emissions (Batterman et al., 2014). These exposures can increase the risk of adverse health consequences, including lung and pulmonary function issues, liver and kidney dysfunction, birth defects (e,g., neurologic impairment), asthma, and respiratory complications (Arif and Shah, 2007; Batterman et al., 2014; Johnson, 1999; Montero-Montoya et al., 2018; Woodruff et al., 2011). While many studies have focused on the effects of individual environmental pollutants, people are often exposed to multi-pollutants and pollution mixtures in real life (Park et al., 2014). To address this gap, recent studies have begun to investigate the effects of multi-pollutants and pollution mixtures (Batterman et al., 2014; Park et al., 2014; Patel et al., 2013).

We analyze the biomonitoring data from the National Health and Nutritional Examination Survey (NHANES) 2003-2012 to evaluate the effects of individual VOC exposures and their two-way interactions on tachycardia (i.e., rapid heart rate) - a risk factor to many cardiovascular diseases - among women of reproductive age. Specifically, we consider four common VOC exposures – Benzene, 1,4-Dichlorobenzene, Ethylbenzene, and methyl tert-butyl ether (MTBE) – for a nationally representative sample of 2003 women aged 15-49 years old in the U.S. population. For more discussions about sources of these VOC exposures, see §S2 in the Supplementary Materials. We are interested in estimating the effects of four binary treatments on whether the VOCs' blood concentration exceeds their detection limits. Our outcome of interest is heart rate. We also have access to people's demographic information (age, race, family income to poverty ratio), hypertension, alcohol consumption, and smoking status. As shown in Table S3 in the Supplementary Materials, the sample sizes and the covariate distributions vary across treatment combination groups.

To remove confounding due to the observed covariates and draw reliable causal conclusions, we employ the proposed weighting estimator with interaction balance constraints and the covariates themselves as basis functions to emulate a 2^4 full factorial design. As shown in Figure S6 in the Supplementary Materials §S2.3, our weighting method greatly improved covariate balances (in terms of absolute standardized mean differences) in all contrasts for main effects and two-way interactions. Successfully controlling for confounding due to age, race, family income to poverty ratio, hypertension, alcohol consumption, and smoking status allows us to estimate the factorial effects of VOC exposures on tachycardia more reliably.

Next, we estimate the factorial effects using our weighting estimator, estimate the variance using the proposed consistent variance estimator, and construct the corresponding 95% confidence intervals, assuming there are no other unmeasured confounders. The results are summarized in Table 1. Our analysis suggests that exposure to 1,4-Dichlorobenzene and MTBE can significantly increase heart rates, but there is no significant evidence for the effects of Benzene and Ethylbenzene. Furthermore, we found no significant evidence of the effects of one VOC on heart rate depending on the other. Our findings highlight the importance of being cautious while selecting home improvement products and choosing living areas with safe air and water quality. Additionally, our results strengthen the reliability of previous studies that have focused on individual chemicals since the interaction effects are insignificant (Arif and Shah, 2007; Montero-Montoya et al., 2018; Weichenthal et al., 2011). In comparison, using regression with treatment-covariate interactions to adjust for the confounding effects (as in the simulation studies) would fail to detect the significant factorial

effects for the four VOC exposures we examined, whereas the weighting method yields more robust and accurate estimates of these effects.

7 Discussion

Although this paper focuses on the factorial effects of binary treatments, the proposed methods can be readily extended to multi-level treatments, albeit with additional notational complexity (see Appendix A of Zhao and Ding (2022)). Specifically, for *l*-level treatments, we redefine A_i^{Ω} over the extended range $\Omega = 1, \ldots, l$ instead of the binary choice $\Omega = +, -$. The balance constraints derived in this paper can then be applied to yield unbiased estimates of factorial effects $\tau_{\mathcal{K}}$. The procedure for estimating weights and conducting inferences for the factorial effects remains the same as discussed in the main text, so we omit the details for brevity.

The weighting framework introduced in this paper relies on the unconfoundedness assumption, which, while commonly adopted in observational studies, may not always hold in practice. A valuable direction for future research would be to investigate how unmeasured confounders might affect the estimation of factorial effects. One possible approach is to conduct sensitivity analyses. While some sensitivity analysis methods have been proposed for the weighting methods (Zhao et al., 2019; Soriano et al., 2023), extending these techniques to accommodate multiple treatments in factorial studies would be a beneficial advance.

Supplementary Material

The supplement file contains additional simulation experiments, further details of the NHANES study, extended discussions and illustrative examples of key concepts from the main paper, and the technical details and proofs of the theoretical results. Code for implementing the proposed method is available at the GitHub repository https://github.com/ruoqiyu/FactorialOS.

Data Availability Statement

Data are publicly available at https://wwwn.cdc.gov/nchs/nhanes/Default.aspx.

Acknowledgements

The authors would like to thank Avi Feller, Luke Keele, Qingyuan Zhao, and José Zubizarreta for insightful discussions.

References

Arif, A. A. and Shah, S. M. (2007). Association between personal exposure to volatile organic compounds and asthma among us adult population. *International Archives of Occupational and Environmental Health*, 80:711–719.

Batterman, S., Su, F.-C., Li, S., Mukherjee, B., and Jia, C. (2014). Personal exposure to mixtures of volatile organic compounds: modeling and further analysis of the riopa data. *Research Report (Health Effects Institute)*, Jun(181):3–63.

Ben-Michael, E., Feller, A., and Hartman, E. (2024). Multilevel calibration weighting for survey data. *Political Analysis*, 32:65–83.

Berenson, G. S., Srinivasan, S. R., Bao, W., Newman, W. P., Tracy, R. E., and Wattigney, W. A. (1998). Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. *New England Journal of Medicine*, 338(23):1650–1656.

Berry, S. M., Connor, J. T., and Lewis, R. J. (2015). The platform trial: an efficient strategy for evaluating multiple treatments. *Journal of the American Medical Association*, 313(16):1619–1620.

Box, G. E., Hunter, W. H., and Hunter, S. (1978). *Statistics for experimenters*. New York: John Wiley and Sons.

Box, G. E. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, 28(1):11–18.

Branson, Z., Dasgupta, T., and Rubin, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study. *The Annals of Applied Statistics*, 10(4):1958–1976.

Bruns-Smith, D. and Feller, A. (2022). Outcome assumptions and duality theory for balancing weights. In *International Conference on Artificial Intelligence and Statistics*, pages 11037–11055. PMLR.

Byar, D. P., Herzberg, A. M., and Tan, W.-Y. (1993). Incomplete factorial designs for randomized clinical trials. *Statistics in Medicine*, 12(17):1629–1641.

Chan, K. C. G., Yam, S. C. P., and Zhang, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B*, 78(3):673–700.

Chattopadhyay, A. and Zubizarreta, J. R. (2022). On the implied weights of linear regression for causal inference. *arXiv preprint arXiv:2104.06581v3*.

Cohn, E. R., Ben-Michael, E., Feller, A., and Zubizarreta, J. R. (2023). Balancing weights for causal inference. In *Handbook of Matching and Weighting Adjustments for Causal Inference*, pages 293–312. Chapman and Hall/CRC.

Dasgupta, T., Pillai, N. S., and Rubin, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *Journal of the Royal Statistical Society: Series B*, 77(4):727–753.

Dong, N. (2015). Using propensity score methods to approximate factorial experimental designs to analyze the relationship between two variables and an outcome. *American Journal of Evaluation*, 36(1):42–66.

Egami, N. and Imai, K. (2019). Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 114(536):529–540:.

Fan, J., Imai, K., Liu, H., Ning, Y., and Yang, X. (2023). Optimal covariate balancing conditions in propensity score estimation. *Journal of Business & Economic Statistics*, 41(1):97–110.

Fong, C. and Imai, K. (2014). Covariate balancing propensity score for general treatment regimes. *Princeton Manuscript*, pages 1–31.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.

Han, S. and Rubin, D. B. (2021). Contrast-specific propensity scores. *Biostatistics & Epidemiology*, 5(1):1–8.

Hirano, K. and Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, 2(3-4):259–278.

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

Imai, K. and Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Johnson, B. L. (1999). A review of the effects of hazardous waste on reproductive health. *American Journal of Obstetrics and Gynecology*, 181(1):S12–S16.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539.

Kuhn, J., Sheldrick, R. C., Broder-Fingert, S., Chu, A., Fortuna, L., Jordan, M., Rubin, D., and Feinberg, E. (2019). Simulation and minimization: technical advances for factorial experiments designed to optimize clinical interventions. *BMC Medical Research Methodology*, 19:1–9.

Li, F. and Li, F. (2019). Propensity score weighting for causal inference with multiple treatments. *The Annals of Applied Statistics*, 13(4):2389–2415.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400.

Li, X., Ding, P., and Rubin, D. B. (2020). Rerandomization in 2^K factorial experiments. *The Annals of Statistics*, 48(1):43–63.

Lopez, M. J. and Gutman, R. (2017). Estimation of causal effects with multiple treatments: a review and new ideas. *Statistical Science*, 32(3):432–454.

Lu, J. (2016). Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statistics & Probability Letters*, 119:11–20.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, 32(19):3388–3414.

Montero-Montoya, R., López-Vargas, R., and Arellano-Aguilar, O. (2018). Volatile organic compounds in air: sources, distribution, exposure and associated illnesses in children. *Annals of Global Health*, 84(2):225.

Mukerjee, R., Dasgupta, T., and Rubin, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association*, 113(522):868–881.

Park, S. K., Tao, Y., Meeker, J. D., Harlow, S. D., and Mukherjee, B. (2014). Environmental risk score as a new tool to examine multi-pollutants in epidemiologic research: an example from the nhanes study using serum lipid levels. *PloS one*, 9(6):e98632.

Pashley, N. E. and Bind, M.-A. C. (2023). Causal inference for multiple treatments using fractional factorial designs. *Canadian Journal of Statistics*, 51(2):444–468.

Patel, C. J., Rehkopf, D. H., Leppert, J. T., Bortz, W. M., Cullen, M. R., Chertow, G. M., and Ioannidis, J. P. (2013). Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the united states national health and nutrition examination survey. *International Journal of Epidemiology*, 42(6):1795–1810.

Resa, M. d. l. A. and Zubizarreta, J. R. (2020). Direct and stable weight adjustment in non-experimental studies with multivalued treatments: analysis of the effect of an earthquake on post-traumatic stress. *Journal of the Royal Statistical Society: Series A*, 183(4):1387–1410.

Rillig, M. C., Ryo, M., Lehmann, A., Aguilar-Trigueros, C. A., Buchert, S., Wulf, A., Iwasaki, A., Roy, J., and Yang, G. (2019). The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science*, 366(6467):886–890.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Robins, J. M., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Rosenbaum, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3):808–840.

Shu, D., Han, P., Hennessy, S., and Miano, T. A. (2023). Robust causal inference of drug-drug interactions. *Statistics in Medicine*, 42(7):970–992.

Soriano, D., Ben-Michael, E., Bickel, P. J., Feller, A., and Pimentel, S. D. (2023). Interpretable sensitivity analysis for balancing weights. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(4):707–721.

Wang, Y. and Zubizarreta, J. R. (2020). Minimal dispersion approximately balancing weights: asymptotic properties and practical considerations. *Biometrika*, 107(1):93–105.

Weichenthal, S., Kulka, R., Dubeau, A., Martin, C., Wang, D., and Dales, R. (2011). Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environmental Health Perspectives*, 119(10):1373–1378.

Wong, R. K. and Chan, K. C. G. (2018). Kernel-based covariate functional balancing for observational studies. *Biometrika*, 105(1):199–213.

Woodruff, T. J., Zota, A. R., and Schwartz, J. M. (2011). Environmental chemicals in pregnant women in the united states: Nhanes 2003–2004. *Environmental Health Perspectives*, 119(6):878–885.

Wu, C. J. and Hamada, M. S. (2021). *Experiments: planning, analysis, and optimization*. New York: John Wiley & Sons.

Yu, R. and Wang, S. (2024). Treatment effects estimation by uniform transformer. *International Conference on Learning Representations*.

Zhao, A. and Ding, P. (2022). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika*, 109(3):799–815.

Zhao, A. and Ding, P. (2023). Covariate adjustment in multiarmed, possibly factorial experiments. *Journal of the Royal Statistical Society: Series B*, 85(1):1–23.

Zhao, A., Ding, P., Mukerjee, R., and Dasgupta, T. (2018). Randomization-based causal inference from split-plot designs. *The Annals of Statistics*, 46(5):1876–1903.

Zhao, Q. (2019). Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, 47(2):965–993.

Zhao, Q., Small, D. S., and Bhattacharya, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(4):735–761.

Zubizarreta, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922.

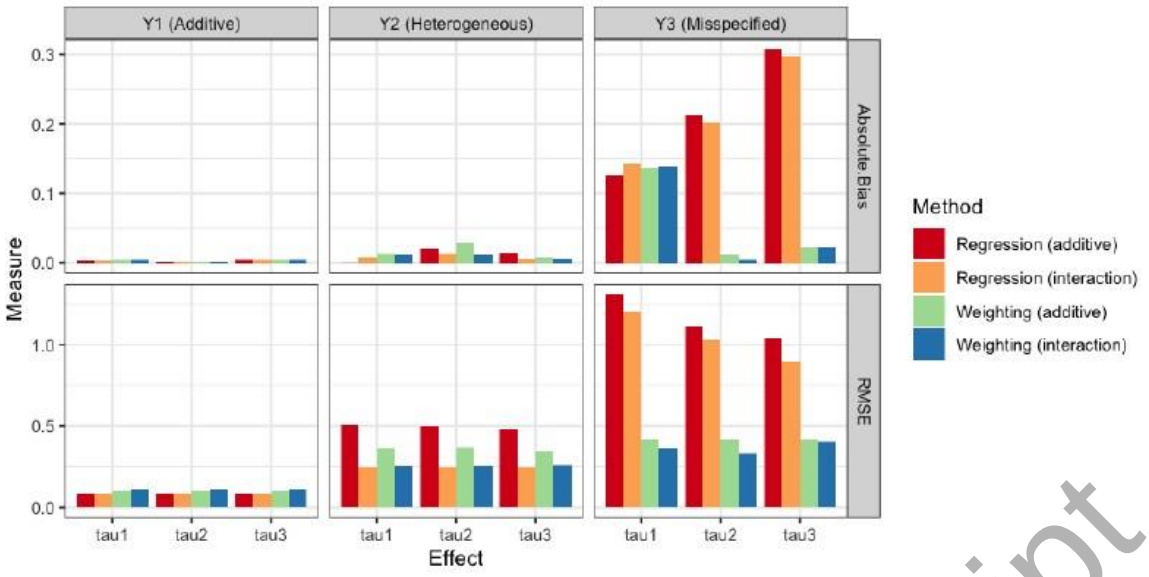

Figure 1: Absolute bias and root mean squared error (RMSE) for estimating three main effects over 1000 repetitions using four estimators when $N = 1000$ and $\rho = 0.4$.

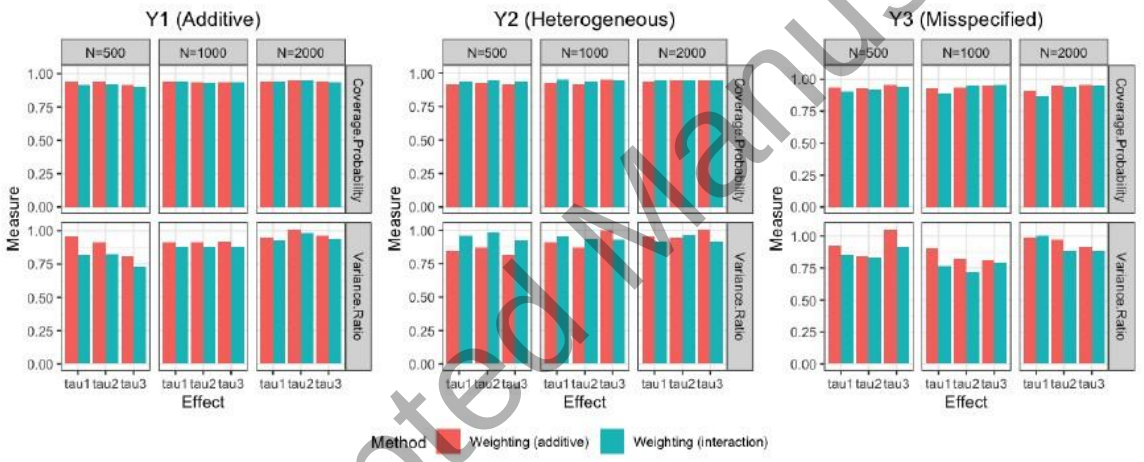

Figure 2: Simulated variance estimator and consistent variance estimator for three main effects ($\sqrt{N}(\hat{\tau}_k - \tau_k), k = 1,2,3$). Variance ratio is the ratio of the consistent variance estimator to the simulated variance estimator. Empirical coverage probabilities of 95% confidence intervals are calculated using the normal distribution and the proposed consistent variance estimator.

Table 1: Factorial effects (four main effects and six second-order interaction effects) for the NHANES study: Estimated factorial effects, consistent estimated variance, corresponding 95% confidence interval (CI). * indicates significant effects at level $\alpha = 0.05$.

	Treatment	Weighting			Regression		
		Estimate	Variance	95% CI	Estimate	Variance	95% CI
Main Effect	Benzene	-0.288	0.643	(-1.860, 1.284)	-0.485	0.917	(-2.362, 1.392)
	1,4-Dichlorobenzene	1.715	0.655	(0.129, 3.300)*	0.320	0.665	(-1.278, 1.918)
	Ethylbenzene	-0.292	0.645	(-1.866, 1.282)	0.016	0.979	(-1.923, 1.955)
	MTBE	2.056	0.658	(0.466, 3.646)*	1.261	0.776	(-0.466, 2.989)
Two-way Interaction	Benzene: 1,4-Dichlorobenzene	0.335	0.644	(-1.238, 1.908)	0.510	0.864	(-1.311, 2.331)
	Benzene: Ethylbenzene	0.992	0.649	(-0.586, 2.571)	1.322	0.898	(-0.535, 3.179)
	Benzene: MTBE	-0.657	0.647	(-2.234, 0.919)	-0.699	0.848	(-2.504, 1.107)
	1,4-Dichlorobenzene: Ethylbenzene	-1.561	0.647	(-3.138, 0.015)	-1.102	0.701	(-2.742, 0.539)
	1,4-Dichlorobenzene: MTBE	-0.183	0.642	(-1.754, 1.387)	-1.127	0.354	(-2.293, 0.040)
	Ethylbenzene: MTBE	-1.556	0.647	(-3.132, 0.020)	-1.029	0.734	(-2.708, 0.650)