Journal of the American Statistical Association

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

Effect Aliasing in Observational Studies

Paul R. Rosenbaum & José R. Zubizarreta

To cite this article: Paul R. Rosenbaum & José R. Zubizarreta (15 Oct 2025): Effect Aliasing in Observational Studies, Journal of the American Statistical Association, DOI: 10.1080/01621459.2025.2537456

To link to this article: https://doi.org/10.1080/01621459.2025.2537456

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 00, NO. 0, 1–12: Theory and Methods
https://doi.org/10.1080/01621459.2025.2537456

Effect Aliasing in Observational Studies

Paul R. Rosenbaum[a] and José R. Zubizarreta[b]

[a]Department of Statistics and Data Science, Wharton School, University of Pennsylvania, Philadelphia, PA; [b]Departments of Health Care Policy, Biostatistics, and Statistics, Harvard University, Boston, MA

ABSTRACT

In experimental design, aliasing of effects occurs in fractional factorial experiments, where certain low order factorial effects are indistinguishable from certain high order interactions: low order contrast weights may be orthogonal to one another, while their higher order interactions are aliased and not identified. In observational studies, aliasing occurs when certain combinations of covariates—for example, time period and various eligibility criteria for treatment—perfectly predict the treatment that an individual will receive, so a covariate combination is aliased with a particular treatment. In this situation, when a contrast among several groups is used to estimate a treatment effect, collections of individuals defined by contrast weights may be balanced with respect to summaries of low-order interactions between covariates and treatments, but necessarily not balanced with respect high-order interactions. We develop a theory of aliasing in observational studies, illustrate that theory in an observational study whose aliasing is more robust than conventional difference-in-differences, and develop a new form of matching to construct balanced confounded factorial designs from observational data. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received June 2024
Accepted July 2025

KEYWORDS
Aliased contrasts; Cardinality matching; Causal inference; Fractional factorial; Incomplete blocks

1. Propensity Scores That Always Equal 0 or 1

In a democracy, when a new policy is adopted by an act of legislation or regulation, it typically applies to everyone who meets explicit criteria while living in a particular region during a particular period of time. Not only is the policy not randomly assigned to individuals, but typically assignment is determined by a few measured covariates, including calendar time, region, and explicit conditions of eligibility. For brevity, we call these "eligibility covariates" and use the symbol $\boldsymbol{w}$ for their values for an individual: they determine treatment assignment, with no stochastic element. Additionally, there are other measured covariates, $\boldsymbol{x}$, that may be associated with group membership, and possibly unmeasured covariates as well. More precisely, individuals fall in several groups, $g = 1, \ldots, G$, and every individual in group g has the same value, say $\boldsymbol{w}_g$, of the eligibility covariates, $\boldsymbol{w}$. Write $Z_g = 1$ if the individual in question is in group g, and $Z_g = 0$ otherwise, so $1 = \sum_{g=1}^{G} Z_g$ for each individual. Because an individual's group is determined by $\boldsymbol{w}$, treatment assignment cannot be ignorable or unconfounded given observed covariates $(\boldsymbol{x}, \boldsymbol{w})$ in the sense of Rosenbaum and Rubin (1983, 2023), because the propensity score, $\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}, \boldsymbol{w}\right)$, is 1 if $\boldsymbol{w} = \boldsymbol{w}_g$, and $\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}, \boldsymbol{w}\right) = 0$ if $\boldsymbol{w} \neq \boldsymbol{w}_g$. There is no common support to the multivariate distributions of measured covariates $(\boldsymbol{x}, \boldsymbol{w})$ in the various groups, because treatment assignment is determined by the observed eligibility covariates, $\boldsymbol{w}$. It may happen that $0 < \Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right) < 1$ for all g, but $\boldsymbol{w}$ cannot simply be ignored.

For example, New Jersey raised its minimum wage from \$4.25 to \$5.05 on April 1, 1992; so, everyone in New Jersey had a \$4.25 minimum before April 1992, and everyone had a minimum of \$5.05 thereafter. The treatment is determined by two measured "eligibility" covariates, namely residence in New Jersey and calendar time. Although one could compare employment levels in New Jersey before and after April 1992 when attempting to estimate the effect of the minimum wage on employment, this comparison is inseparable from changes in general economic conditions from before to after April 1, 1992. By analogy with experimental design, one might say that in a comparison confined to New Jersey, the main effect of the change in the minimum wage is aliased with the time period. In a widely discussed study, Card and Krueger (1994) compared the change in employment at fast-food chains like Burger King in New Jersey to the change at the same chains in the adjacent state of Pennsylvania where the minimum wage remained at \$4.25 in the periods before and after April 1, 1992. The study attracted attention among economists because it found no indication of a decline in employment caused by the minimum wage, and it therefore stood opposed to a long standing prediction based on economic theory (Stigler 1946).

A comparison of this sort is often called the "method of difference-in-differences" in which the time-by-state interaction is taken as an estimate of the effect of the minimum wage. It is widely recognized that this interaction estimate is not biased by an additive temporal shift in employment that equally affects New Jersey and Pennsylvania, nor by an additive shift in employment that distinguishes New Jersey and Pennsylvania in the

CONTACT José R. Zubizarreta zubizarreta@hcp.med.harvard.edu Departments of Health Care Policy, Biostatistics, and Statistics, Harvard University, 180 A Longwood Ave, Office 215-A, Boston, MA 02115.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

same way in both time periods, but it could easily be biased in many other ways, not least by a temporal trend in New Jersey that differs from the trend in Pennsylvania; see Rosenbaum (2021, sec. 4.4) for additional strengths of this design. Informally borrowing terminology from experimental design, the effect of the treatment is aliased with, or indistinguishable from, the interaction of state and time, but it is not aliased with the main effect of state or the main effect of time. Stated informally in a different way that we develop formally in Definition 2 in Section 3.1, if employment levels were distorted by an additive bias, $\beta(\boldsymbol{x}, \boldsymbol{w})$, that is the sum of two terms, one depending upon $\boldsymbol{x}$ and only the state in $\boldsymbol{w}$, the other depending upon $\boldsymbol{x}$ and only the time period in $\boldsymbol{w}$, then that two-term bias would cancel out when the difference-in-differences or interaction contrast is calculated. We would know that $\beta(\boldsymbol{x}, \boldsymbol{w})$ cancels if it had this specific two-term form, without having to estimate $\beta(\boldsymbol{x}, \boldsymbol{w})$, providing we had appropriately matched for $\boldsymbol{x}$. Matching for $(\boldsymbol{x}, \boldsymbol{w})$ is not possible because the groups do not overlap on $\boldsymbol{w}$. For various perspectives on difference-in-differences, see Cook, Campbell, and Shadish (2002), Meyer (1995), Shaikh and Toulis (2021) and Ye et al. (2024).

In the literature on difference-in-differences designs, the state-by-time interaction contrast is commonly taken as the obvious estimate of the effect of raising the minimum wage in the later time period, having removed the state and time period main effects. The same contrast can be derived, instead of taken as obvious, by: (i) building a 4-row design matrix with a first column of 1's for a constant term, a column $(-1, -1, 1, 1)$ for before and after, a column $(-1, 1, -1, 1)$ for Pennsylvania or New Jersey, and a column for $(0, 0, 0, 1)$ for the increase in the minimum wage in New Jersey in the after period, (ii) orthogonalizing the fourth column to the previous three columns, (iii) thereby producing a multiple of the difference-in-differences contrast $(1, -1, -1, 1)$ in place of $(0, 0, 0, 1)$. Although this and other similar derivations work quite generally, we follow the literature and take difference-in-differences contrasts as obvious and do not derive them.

The difference-in-differences design is analogous to a half-fraction of a 2^3 factorial design, in which main effects of time, state and policy are not aliased with each other, but each main effect is aliased with the two-factor interaction of the other two factors. This analogy is limited because it ignores various facts that we address later: (i) time and state are covariates, and only the policy is a treatment, (ii) there is abundant data from this half-fraction and no data at all from the complementary half fraction, (iii) because individuals are not randomly assigned to groups, there can be additional measured and unmeasured confounding by covariates not captured by eligibility covariates, (iv) aliasing is imposed upon the investigator by the manner in which the legislature implemented its policy. Wu (2015) wrote: "Effect aliasing is a basic concept and necessary evil in fractional factorial design," and one might say the same about effect aliasing in policy evaluations with eligibility covariates that determine treatment assignment.

The interaction contrast or difference-in-differences is a contrast of several groups of people having the property that it is not aliased or confused with certain simple contrasts of bias terms, but it is aliased with certain more complex contrasts of bias terms. We formalize this as a weighted sum of bias terms, $\beta(\boldsymbol{x}, \boldsymbol{w})$, say $\beta(\boldsymbol{x}, \boldsymbol{w}) = \sum_{m=1}^{M} \lambda_m \delta_m(\boldsymbol{x}, \boldsymbol{w})$, where the $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ have a specified form and the λ_m are unspecified and unknown constants. In the Card and Krueger example, $\boldsymbol{w} = (w_1, w_2)$ where w_1 is ± 1 distinguishing before or after April 1, 1992, and w_2 is ± 1 distinguishing New Jersey and Pennsylvania. Then $\beta(\boldsymbol{x}, \boldsymbol{w})$ as a weighted sum of bias terms will be an unproblematic bias, after matching to balance $\boldsymbol{x}$, if the terms $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ in the weighted sum include: (i) a term, say $\delta_1(w_1) = w_1$, that only depends on time, w_1; (ii) a term, say $\delta_2(w_2) = w_2$, that depends only on the state, w_2; (iii) a term, say $\delta_3(x_1) = x_1$ that depends on the employee's sex, say x_1; (iv) an interaction term, say $\delta_4(w_2, x_1) = w_2 \cdot x_1$ that depends jointly on the state w_2 and the employee's sex, x_1; (v) an interaction term, say $\delta_5(w_1, x_1, x_2, x_3) = w_1 \cdot x_1 \cdot x_2 \cdot x_3$, that depends jointly on before-or-after 1 April 1992, w_1, the employee's sex, x_1, the employee's age, x_2, and a binary indicator that the specific fast food chain is Burger King, say x_3. As discussed in Section 3, balancing $\boldsymbol{x}$ by matching and calculating the difference-in-differences contrast will annihilate such a bias term, $\beta(\boldsymbol{x}, \boldsymbol{w}) = \sum_{m=1}^{M} \lambda_m \delta_m(\boldsymbol{x}, \boldsymbol{w})$, even though $(\boldsymbol{x}, \boldsymbol{w})$ perfectly predicts treatment, with a 0-or-1 propensity score. Moreover, we can check that the matching has succeeded in this limited task by checking for balance when the contrast is applied to each of the M basis functions, $\delta_m(\boldsymbol{x}, \boldsymbol{w})$; see Section 3.3. However, the sum $\beta(\boldsymbol{x}, \boldsymbol{w})$ would be irreparably problematic if it contains an additive term that depends in a nontrivial way on (w_1, w_2) jointly, such as $w_1 \cdot w_2$. See Proposition 1 for the first and simplest result of this kind. Viewed in this way, we have reason to prefer designs that resist more complicated bias terms, $\beta(\boldsymbol{x}, \boldsymbol{w})$, and reason to avoid designs that are irreparably damaged by simple bias terms.

To illustrate, beginning in Section 2, we use an interesting study by Lalive, Van Ours, and Zweimüller (2006) that has a more complex, more robust alias structure than does the usual difference-in-differences design. To describe the alias structure as more robust is to say that certain bias structures, $\beta(\boldsymbol{x}, \boldsymbol{w})$, that would bias the difference-in-differences design do not bias this more robust alias structure. Because group assignment is determined by $\boldsymbol{w}$, a sufficiently complex bias, $\beta(\boldsymbol{x}, \boldsymbol{w})$, would invalidate both designs, and we explore this as well. Moreover, different treatment contrasts within a single study resist biases of different complexity, and we ultimately prefer the conditional main effect contrasts proposed by Wu (2015), Su and Wu (2017), and Wu and Hamada (2021, secs. 4.3.2, 5.5).

2. An Observational Study of Unemployment Benefits

2.1. Treatments Aliased with Eligibility × Time Interactions

In August 1989, Austria changed its unemployment benefits, increasing the monetary benefit amount (the replacement rate) for workers whose previous jobs had low earnings (LE), and increasing the duration of benefits for older workers with a long period of no or infrequent unemployment (IU). An increase in benefits duration is denoted B and no increase is b, while an increase in the replacement rate is R and no increase is r. Lalive, Van Ours, and Zweimüller (2006) studied the effects of these changes on the duration of unemployment. They studied people who became unemployed in the two years before August 1989

Table 1. Contrast matrix among 8 groups, $g = 1, \ldots, G = 8$. Factor Ti is Time.

Group		Treatment		Eligibility			Conditional Main Effects and Block Types, 1 to 6					
							1	2	3	4	5	6
		Bb	Rr	LE w'	IU w''	Ti w'''	Rr@B	Bb@R	BR-vs -br	Br-vs bR	Bb@r	Rr@b
	g	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
BR	1	1	1	1	1	1	1	1	1	0	0	0
Br	2	1	-1	-1	1	1	-1	0	0	1	1	0
bR	3	-1	1	1	-1	1	0	-1	0	-1	0	1
br	4	-1	-1	-1	-1	1	0	0	-1	0	-1	-1
$\overline{\text{BR}}$	5	-1	-1	1	1	-1	-1	-1	-1	0	0	0
$\overline{\text{Br}}$	6	-1	1	-1	1	-1	1	0	0	-1	-1	0
$\overline{\text{bR}}$	7	1	-1	1	-1	-1	0	1	0	1	0	-1
$\overline{\text{br}}$	8	1	1	-1	-1	-1	0	0	1	0	1	1

and other people who became unemployed in the two years after August 1989. Their study is discussed in a textbook by Cahuc, Carcillo, and Zylberberg (2014, chap. 5).

We focus on individuals aged 40–55, as no one under 40 was eligible for a lengthened duration of benefits. The categories, LE and IU, exist before and after August 1989, but they determine B-or-b and R-or-r only after August 1989. After August 1989, IU determines B-or-b, and LE determines R-or-r. Unlike a conventional difference-in-differences design, an interaction between Low Earnings (LE) and Time does not lead to aliasing of the B-or-b effect, and an interaction between Infrequent Unemployment (IU) and Time does not lead to aliasing of the R-or-r effect.

There are four treatment types: BR for an increase in benefit amount and duration, Br for an increase in benefit duration only, bR for an increase in benefit amount only, and br for no increase in benefits. Before August 1989, these four eligibility categories exist, but the changes had not gone into effect; so the categories are denoted $\overline{\text{BR}}$, $\overline{\text{Br}}$, $\overline{\text{bR}}$, $\overline{\text{br}}$ in the period before August 1989. Table 1 contains the relevant contrasts. Contrasts (1) and (2) are, respectively, the difference-in-differences calculations for B-versus-b and for R-versus-r. Contrasts (3), (4), and (5) represent the eligibility covariates, (3) for low earnings or LE, (4) for infrequent unemployment or IU, and (5) for time period. The five main effect contrasts, (1)–(5), in Table 1 are orthogonal to one another; however, they are aliased with certain two-factor interactions. As is commonly true in difference-in-differences comparisons, the interaction or product of columns (4) and (5) equals column (1), because covariates (w'', w''') determine eligibility for an increase in benefit duration, B/b. So, the main effect of B/b is aliased with the interaction of (w'', w'''). In parallel, as is commonly true, the interaction of product of columns (3) and (5) equals column (2), because covariates (w', w''') determine eligibility for an increase in the replacement rate, R/r.

Unlike the usual difference-in-differences design—that is, improving upon the usual difference-in-differences design—some two-factor interactions are not aliased with main effects, but rather with other two-factor interactions. Table 2 shows the aliasing of the main effects and two-factor interactions in Table 1. Because main effects in Table 1 are not aliased with each other, the columns of Table 2 include only interactions. A "■" signifies two totally aliased contrasts; that is, the contrasts have different names but are numerically identical. A blank in Table 2 signifies orthogonal contrasts. There are eight groups of people

Table 2. Aliasing of main effects and 2-factor interactions in Table 1.

Effects	Two-factor Interactions									
	Bb×Rr	Bb×LE	Bb×IU	Rr×LE	Rr×IU	LE×IU	Ti×Bb	Ti×Rr	Ti×LE	Ti×IU
Ti			■	■						
Bb										■
Rr									■	
LE								■		
IU							■			
Bb×Rr						■				
Bb×LE					■					
Bb×IU				■						
Rr×LE			■							
Rr×IU		■								
LE×IU	■									
Ti×Bb										
Ti×Rr										
Ti×LE										
Ti×IU										

NOTE: A "■" signifies totally aliased contrasts; a blank indicates orthogonal contrasts. Ti is Time.

in Table 1, and hence $8 - 1 = 7$ linearly independent contrasts among the eight groups. There are five orthogonal main effect contrasts, (1)–(5), in Table 1, and $(5 \times 4)/2 = 10$ two-factor interactions or products of two of these contrasts, making $15 = 5 + 10$ main effects plus two factor interactions. In Table 2, three of the 15 effects are the same as the time period contrast (namely Ti, Bb×IU and Rr×LE), and the remaining $15-3 = 12$ contrasts are aliased in 6 pairs, where $15 = 1 \times 3 + 6 \times 2$. For instance, the change in benefits duration, B/b, is determined by a history of infrequent unemployment, IU, while an increase in money paid, R/r, is determined low earnings, LE, in the job just lost; so, it is pleasant news in Table 2 that neither the $B/b \times LE$ interaction, nor the $R/r \times IU$ interaction is aliased with any main effect; however, these two interactions are aliased with each other. Also evident in Table 2 is the disappointing fact that the $B/b \times R/r$ interaction is directly aliased with the $IU \times LE$ interaction; that is, having two time periods dealiases the main effects of B/b and R/r from the main effects of IU and LE, but it fails to dealias their two-factor interactions.

Subtleties are in Web-appendix Table 1. The benefits available prior to August 1989 depended upon whether one has age ≤ 50 and whether one worked for at least 3 of the prior 5 years. These two covariates do not determine eligibility for a benefits change in August 1989—for instance, in every group g in Table 3, there are people older and younger than 50; so, these two covariates belong to $\boldsymbol{x}$, not $\boldsymbol{w}$. It is important to balance these covariates; see Sections 3.3, 4.1, and 6.3.

2.2. Eight Treatments in Blocks of Size 4

As seen in Table 3, our blocked comparison contains 16,800 individuals in 4200 blocks of size 4, where $16{,}800 = 4200 \times 4$. The matching that formed the blocks of size 4 is described in Section 5. There are six types of blocks, with 700 blocks of each type, or $2800 = 700 \times 4$ individuals in each block type. Each of the six block types is one of the six possible difference-in-differences comparisons of two of the four treatments combinations. These difference-in-differences effects in the several block types closely resemble the "conditional main effects" of Wu (2015), Su and Wu (2017), and Wu and Hamada (2021, secs. 4.3.2, 5.5), in which

Table 3. Six block types, 1–6, each containing two eligibility categories $\overline{\text{Before}}$ and After two treatments were applied in the After period.

		Eight treatments in 6 types of blocks of size 4						N
Time	h_g	1 BR/Br	2 BR/bR	3 BR/br	4 Br/bR	5 Br/br	6 bR/br	
After	1	BR	BR	BR	Br	Br	bR	4200
After	−1	Br	bR	br	bR	br	br	4200
$\overline{\text{Before}}$	−1	$\overline{BR}$	$\overline{BR}$	$\overline{BR}$	$\overline{Br}$	$\overline{Br}$	$\overline{bR}$	4200
$\overline{\text{Before}}$	1	$\overline{Br}$	$\overline{bR}$	$\overline{br}$	$\overline{bR}$	$\overline{br}$	$\overline{br}$	4200
N		2800	2800	2800	2800	2800	2800	16,800

NOTE: Each block type occurs 700 times, so there are $4 \times 700 = 2800$ individuals in a column, $6 \times 700 = 4200$ in a row, and $2800 \times 6 = 16{,}800$ in the study. Note that block type 1 is confined to the population with infrequent unemployment (IU), while type 6 is confined to the population with more frequent unemployment. Also, block type 2 is confined to the population with low earnings (LE), while type 5 is confined to the population with higher earnings.

Table 4. Relationships in Table 1 between traditional main effects and two factor interaction contrasts and the conditional main effects in the six types of blocks. A blank indicates orthogonal contrasts. A "•" indicates contrasts that are not orthogonal but are logically related, such as a main effect and a conditional main effect of the same factor. A "■" indicates non-orthogonal contrasts that are not logically related. All non-orthogonal contrasts have squared correlation of 1/2.

	Block Type and Conditional Main Effect					
Block Type	1	2	3	4	5	6
Effect	Rr@B	Bb@R	BR-vs-br	Br-vs-bR	Bb@r	Rr@b
Bb		•	•	•	•	
Rr	•		•	•		•
LE						
IU						
Time						
Bb×Rr						
Bb×LE	■	■			■	■
Bb×IU						
Rr×LE						
Rr×IU	■	■			■	■
LE×IU						
Time×Bb						
Time×Rr						
Time×LE	■		■	■		■
Time×IU		■	■	■	■	

one studies the effect of one factor at a specific level of another. For a discussion of confounding of factorial effects in incomplete blocks as it occurs in experiments, see Bailey (2008, chap. 12) and Wu and Hamada (2021, secs. 4.15, 5.7, 6.8).

Block type 2 compares the high and low levels, B and b, of the change in benefits duration, for individuals with low earnings, LE, who would receive the high level increase, R, in the replacement rate. Blocks of type 2 are homogeneous in having low earnings in their prior job. In Table 1, this is a comparison of groups $g = 1, 3, 5$, and 7, with difference-in-differences contrast $(BR - bR) - \left(\overline{BR} - \overline{bR}\right)$, or the equivalent 8-group contrast $Bb@R$ in column (7); moreover, $Bb@R$ is orthogonal to the corresponding four rows of main-effects in columns (2), (3), (4), and (5). The contrast $(BR - bR) - \left(\overline{BR} - \overline{bR}\right)$ picks out rows of Table 1 with a +1 in column (3). Here, $Bb@R$ means the effect of B-vs-b at level R of Rr. In block type 5 in Table 3 is B-versus-b for individuals ineligible for an increase in the replacement rate, r; it is $Bb@r$ in column (10) in Table 1.

Block types 1 and 6 examine R-versus-r at specific levels of B/b, thereby reversing the roles of B/b and R/r. Consequently, these blocks are homogeneous in having infrequent periods of unemployment in the past, IU. Block types 1, 2, 5, and 6 are homogeneous in one eligibility covariate, IU or LE; so, these types refer to populations defined by a covariate.

In block type 3, both enhancements to benefits are compared to neither enhancement, namely BR versus br, in contrast $(BR - br) - \left(\overline{BR} - \overline{br}\right)$ among rows 1, 4, 5, and 8 of Table 1. This is expressed as column (8) of Table 1, which is orthogonal to the main effects (3), (4) and (5). In blocks of type 4, each enhancement is compared in lieu of the other, Br versus bR, as expressed in column (9). Neither LE nor IU can be homogeneous in block types 3 and 4.

There are eight groups in Table 1; so, one might consider a design with blocks of size 8 representing each group once; however, blocks of size four can be, and have been, matched to be more homogeneous than blocks of size eight could possibly be. For example, in a block of type 2, eligibility covariate LE or w' is perfectly matched with $w' = 1$, but this is impossible in blocks of size eight because groups Br, br, $\overline{Br}$, and $\overline{br}$ always have $w' = -1$. Similarly, blocks that estimate conditional main effects of R-vs-r can be, and have been, matched to be homogeneous in IU or w'', but this is impossible in blocks that include all eight groups. Blocks of types 3 and 4 vary both B/b and R/r; so, they can be matched for neither w' nor w''.

In experiments, reducing heterogeneity reduces the standard error of an unbiased estimate of effect. Reducing within-block heterogeneity is more important in observational studies, because it removes a bias from measured covariates, and by virtue of reducing heterogeneity it increases insensitivity to unmeasured biases (Rosenbaum 2005; Zubizarreta, Paredes, and Rosenbaum 2014).

The six block types in the block design in Table 3 permit six separate estimates of six conditional effects. Table 4 relates these six effects to the traditional main effects and two-factor interactions in Table 2. By definition the sum of the conditional main effect $Bb@R$ in block type 2 and $Bb@r$ in block type 5 is the main effect of Bb in Table 2, and the difference of these conditional main effects is the $Bb \times Rr$ interaction; so, these effects are logically related and not orthogonal in Table 4, as indicated by a "•". Table 4 also notes by a "■" the partial aliasing of effects created by fractional replication in the design imposed upon Lalive, Van Ours, and Zweimüller (2006) by the structure of Austria's benefit reform. Notably in Table 4: (i) Time×Bb and Time×Rr are aliased with main effects in Table 2, but they are orthogonal to all of the conditional effects in Table 4; (ii) neither BR-vs-br in block type 3 nor Br-vs-bR in block type 4 is aliased with the LE×IU interaction, unlike several other effects in Table 2 that jointly involve Bb and Rr.

3. Basic Theory of Aliasing in Observational Studies

3.1. Ignorable Treatment Assignment with Aliasing

There are G groups, $g = 1, \ldots, G$, where $\boldsymbol{Z} = (Z_1, \ldots, Z_G)$ indicates the group to which an individual is assigned, with $Z_g = 1$ if the individual is in group g, and $Z_g = 0$ otherwise. Because each individual is assigned to one group, $1 = \sum_{g=1}^{G} Z_g$. In Table 1, $G = 8$. Two types of observed covariates are distinguished. By definition, "aliased eligibility covariates," $\boldsymbol{w}$, determine the group to which an individual is assigned. In Card

and Krueger's (1994) study in Section 1, the aliased eligibility covariates $\boldsymbol{w}$ record the state, New Jersey or Pennsylvania, and the two time periods. The remaining observed covariates, $\boldsymbol{x}$, may be associated with and predictive of treatment assignment when unaided by $\boldsymbol{w}$, but the prediction is imperfect, so two people with the same $\boldsymbol{x}$ might easily receive different treatments. When needed for clarity, $\boldsymbol{x}$ will be called "unaliased covariates." In Table 3, $\boldsymbol{w} = \left(w', w'', w'''\right)$ where w' is low income (LE), w'' is infrequent unemployment (IU) and w''' is time, before or after the change in policy. Group membership $\boldsymbol{Z}$ is determined by $\boldsymbol{w}$, and conversely everyone with $Z_g = 1$ has the same value, say $\boldsymbol{w}_g$, of the eligibility covariates; that is, $\Pr\left(Z_g = 1 \mid \boldsymbol{x},\ \boldsymbol{w} = \boldsymbol{w}_g\right) = 1$ and $\Pr\left(Z_g = 1 \mid \boldsymbol{x},\ \boldsymbol{w} \neq \boldsymbol{w}_g\right) = 0$ for $g = 1, \ldots, G$. Note that $\boldsymbol{w}$ is a random variable, while $\boldsymbol{w}_g$ is one of the G possible values of $\boldsymbol{w}$.

An individual has response r_g if given treatment g, but the effect $r_g - r_{g'}$ caused by receiving treatment g in lieu of treatment g' is not observed for any individual. The observed response is $R = \sum_{g=1}^{G} Z_g\, r_g$. Dawid (1979, Thm. 3) writes $A \perp\!\!\!\perp B \mid C$ for A is conditionally independent of B given C and notes that $A \perp\!\!\!\perp B \mid C$ implies $B \perp\!\!\!\perp A \mid C$.

Definition 1. Treatment assignment is ignorable given $\boldsymbol{x}$ with aliasing by $\beta(\cdot, \cdot)$ if

$$\{r_1 - \beta(\boldsymbol{x}, \boldsymbol{w}), \ldots, r_G - \beta(\boldsymbol{x}, \boldsymbol{w})\} \perp\!\!\!\perp \boldsymbol{Z} \mid \boldsymbol{x}, \tag{1}$$

$$0 < \Pr\left(Z_g = 1 \mid \boldsymbol{x}\right) < 1 \text{ for } g = 1, \ldots, G. \tag{2}$$

In Definition 1, $\Pr\left(Z_g = 1 \mid \boldsymbol{x}, \boldsymbol{w}\right)$ may be 0 or 1, but $\Pr\left(Z_g = 1 \mid \boldsymbol{x}\right)$ is never 0 or 1.

If (2) is true, then the function $\mathrm{E}\left(R \mid Z_g = 1, \boldsymbol{x}\right)$ is estimable as it involves only quantities that are jointly observed: it is the regression of R on $\boldsymbol{x}$ in the nonempty subpopulation with $Z_g = 1$. If $0 = \beta(\boldsymbol{x}, \boldsymbol{w}_1) = \cdots = \beta(\boldsymbol{x}, \boldsymbol{w}_G)$, then Definition 1 is the same as ignorable treatment assignment given $\boldsymbol{x}$ alone in Rosenbaum and Rubin (1983); so, in this special case, $\mathrm{E}\left(r_g - r_{g'} \mid \boldsymbol{x}\right)$ equals

$$\begin{aligned} &\mathrm{E}\left(r_g \mid Z_g = 1, \boldsymbol{x}\right) - \mathrm{E}\left(r_{g'} \mid Z_{g'} = 1, \boldsymbol{x}\right) \\ &\quad = \mathrm{E}\left(R \mid Z_g = 1, \boldsymbol{x}\right) - \mathrm{E}\left(R \mid Z_{g'} = 1, \boldsymbol{x}\right); \end{aligned} \tag{3}$$

so, the expected causal effect, $\mathrm{E}\left(r_g - r_{g'} \mid \boldsymbol{x}\right)$ can be estimated from observed quantities. Of course, (3) used $0 = \beta(\boldsymbol{x}, \boldsymbol{w}_1) = \cdots = \beta(\boldsymbol{x}, \boldsymbol{w}_G)$ and does not otherwise follow from (1). In brief, Definition 1 states a weaker condition than ignorable treatment assignment given $\boldsymbol{x}$, in the sense that the value of $\boldsymbol{Z}$ is a 1-to-1 function of $\boldsymbol{w}$.

Let $h_g, g = 1, \ldots, G$, be constants with $0 = \sum_{g=1}^{G} h_g$ and at least one h_g is not zero. Mukerjee, Dasgupta, and Rubin (2018) discuss contrasts for causal effects in randomized experiments. Suppose that we wish to estimate the expectation of a specific causal contrast, $\mathrm{E}\left(\sum_{g=1}^{G} h_g\, r_g \,\middle|\, \boldsymbol{x}\right) = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid \boldsymbol{x}\right)$. In (3), for instance, $\mathrm{E}\left(r_g - r_{g'} \mid \boldsymbol{x}\right) = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid \boldsymbol{x}\right)$ with $h_g = 1 = -h_{g'}$ and $h_{g''} = 0$ for $g'' \notin \{g, g'\}$. Alternatively, h_g might be given by any one of columns (1), (2), (6)–(11) in Table 1.

Definition 2. Contrast h_g is not aliased with $\beta(\boldsymbol{x}, \boldsymbol{w})$ if $\sum_{g=1}^{G} h_g\, \beta\left(\boldsymbol{x}, \boldsymbol{w}_g\right) = 0$ for all $\boldsymbol{x}$.

For example, if

$$\begin{aligned} \beta(\boldsymbol{x}, \boldsymbol{w}) &= \sum_{m=1}^{4} \lambda_m\, \delta_m(\boldsymbol{x}, \boldsymbol{w}) \\ &= \lambda_1\, w' + \lambda_2\, w'' + \lambda_3\, w'\, w''' + \lambda_4\, x_1\, w', \end{aligned} \tag{4}$$

then for every possible value of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, the contrast h_g, $g = 1, \ldots, G$ in column (7) of Table 1 is not aliased with $\beta(\boldsymbol{x}, \boldsymbol{w})$, as is seen from column 2 of Table 4. However, $\beta(\boldsymbol{x}, \boldsymbol{w})$ would be aliased with column (7) of Table 1 if $\lambda_5 w'' w'''$ were added to (4); see, again, column 2 of Table 4, noting the black squares.

Proposition 1. If treatment assignment is ignorable given covariates $\boldsymbol{x}$ with aliasing by functions $\beta(\boldsymbol{x}, \boldsymbol{w})$, and if contrast h_g, $g = 1, \ldots, G$, is not aliased with $\beta(\boldsymbol{x}, \boldsymbol{w})$, then the expected causal contrast $\mathrm{E}\left(\sum_{g=1}^{G} h_g\, r_g \,\middle|\, \boldsymbol{x}\right)$ equals a contrast of estimable expectations

$$\mathrm{E}\left(\sum_{g=1}^{G} h_g\, r_g \,\middle|\, \boldsymbol{x}\right) = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(R \mid Z_g = 1, \boldsymbol{x}\right).$$

Proof. Because (2) holds, $\mathrm{E}\left(R \mid Z_g = 1, \boldsymbol{x}\right)$ is estimable from observable joint distributions for $g = 1, \ldots, G$. Using the two equalities, $0 = \sum_{g=1}^{G} h_g\, \beta\left(\boldsymbol{x}, \boldsymbol{w}_g\right)$ and $0 = \sum_{g=1}^{G} h_g$, we have:

$$\begin{aligned} &\sum_{g=1}^{G} h_g\, \mathrm{E}\left(R \mid Z_g = 1, \boldsymbol{x}\right) \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid Z_g = 1, \boldsymbol{x}\right), \text{ as } R = r_g \text{ if } Z_g = 1 \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid Z_g = 1, \boldsymbol{x}\right) - \sum_{g=1}^{G} h_g\, \beta\left(\boldsymbol{x}, \boldsymbol{w}_g\right) \text{ as} \\ &\qquad\qquad 0 = \sum_{g=1}^{G} h_g\, \beta\left(\boldsymbol{x}, \boldsymbol{w}_g\right) \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left\{r_g - \beta(\boldsymbol{x}, \boldsymbol{w}) \mid Z_g = 1, \boldsymbol{x}\right\} \text{ as} \\ &\qquad\qquad \boldsymbol{w} = \boldsymbol{w}_g \text{ iff } Z_g = 1 \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left\{r_g - \beta(\boldsymbol{x}, \boldsymbol{w}) \mid \boldsymbol{x}\right\} \text{ by (1)} \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid \boldsymbol{x}\right) - \mathrm{E}\left\{\beta(\boldsymbol{x}, \boldsymbol{w}) \mid \boldsymbol{x}\right\} \sum_{g=1}^{G} h_g \\ &\quad = \sum_{g=1}^{G} h_g\, \mathrm{E}\left(r_g \mid \boldsymbol{x}\right) \text{ because } 0 = \sum_{g=1}^{G} h_g. \end{aligned}$$

□

3.2. Propensity Scores and Balancing Scores

Much of the basic theory of ignorable treatment assignment and propensity scores (Rosenbaum and Rubin 1983, 2023) carries over to the situation with aliasing, even when treatment assignment is a deterministic function of the aliased covariates, $\boldsymbol{w}$, as we briefly demonstrate. The practical point of Proposition 3 is that Proposition 1 is useful without conditioning on all $\boldsymbol{x}$; rather, it suffices to condition or match on enough to balance $\boldsymbol{x}$.

Define $\boldsymbol{e}(\boldsymbol{x}) = \{\Pr(Z_1 = 1 \mid \boldsymbol{x}), \ldots, \Pr(Z_G = 1 \mid \boldsymbol{x})\} = \{e_1(\boldsymbol{x}), \ldots, e_G(\boldsymbol{x})\}$. Proposition 2 is due to Imai and Van Dyk (2004) and does not involve aliasing.

Proposition 2. (Imai and Van Dyk) (i) $\boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{x} \mid \boldsymbol{e}(\boldsymbol{x})$ and

$$\text{(ii) } \boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{x} \mid \left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\} \text{ for any } \boldsymbol{f}(\cdot). \tag{5}$$

Proof. To demonstrate (i) the task is to show $\Pr(\boldsymbol{Z} \mid \boldsymbol{x}) = \Pr\{\boldsymbol{Z} \mid \boldsymbol{e}(\boldsymbol{x})\}$, which follows from:

$$\begin{aligned}\Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{e}(\boldsymbol{x})\right\} &= \mathrm{E}\left[\Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{x}\right\} \mid \boldsymbol{e}(\boldsymbol{x})\right] \\ &= \mathrm{E}\left\{e_g(\boldsymbol{x}) \mid \boldsymbol{e}(\boldsymbol{x})\right\} \\ &= e_g(\boldsymbol{x}) = \Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right).\end{aligned}$$

Also, (ii) follows from (i) by Lemma 4.2 of Dawid (1979). □

Although Proposition 1 gives a sufficient condition for the estimability of a causal contrast, $\mathrm{E}\left(\sum_{g=1}^{G} h_g r_g \,\middle|\, \boldsymbol{x}\right)$, a weaker condition is given in Proposition 3. Instead of matching or otherwise adjusting for all of $\boldsymbol{x}$, it suffices to adjust for a function of $\boldsymbol{x}$, specifically for any function that is at least as fine as the propensity score, that is for any $\left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$. Taking $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ in Proposition 3 yields Proposition 1, whereas if $\boldsymbol{f}(\boldsymbol{x})$ is set to a constant that ignores $\boldsymbol{x}$, then Proposition 3 says it suffices to match or block for $\boldsymbol{e}(\boldsymbol{x})$ alone. Proposition 3 says that we need not match groups exactly for $\boldsymbol{x}$, providing $\boldsymbol{x}$ is balanced across groups in matched blocks in the sense that (5) holds.

Proposition 3. If treatment assignment is ignorable given covariates $\boldsymbol{x}$ with aliasing by functions $\beta(\boldsymbol{x}, \boldsymbol{w})$, then for any $\boldsymbol{f}(\cdot)$ it is also ignorable given $\left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$ with aliasing by functions $\beta(\boldsymbol{x}, \boldsymbol{w})$, that is,

$$\{r_1 - \beta(\boldsymbol{x}, \boldsymbol{w}), \ldots, r_G - \beta(\boldsymbol{x}, \boldsymbol{w})\} \perp\!\!\!\perp \boldsymbol{Z} \mid \left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}, \tag{6}$$

$$0 < \Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\} < 1 \text{ for } g = 1, \ldots, G; \tag{7}$$

so, the expected causal contrast $\mathrm{E}\left\{\sum_{g=1}^{G} h_g r_g \,\middle|\, \boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$ equals a contrast of estimable expectations

$$\begin{aligned}&\mathrm{E}\left\{\sum_{g=1}^{G} h_g r_g \,\middle|\, \boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\} \\ &\quad= \sum_{g=1}^{G} h_g \,\mathrm{E}\left\{R \mid Z_g = 1, \boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}.\end{aligned} \tag{8}$$

Proof. Write $\boldsymbol{d} = \{r_1 - \beta(\boldsymbol{x}, \boldsymbol{w}), \ldots, r_G - \beta(\boldsymbol{x}, \boldsymbol{w})\}$ and $\boldsymbol{s}(\boldsymbol{x}) = \left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$. Because $A \perp\!\!\!\perp B \mid C$ implies $B \perp\!\!\!\perp A \mid C$, to demonstrate (6) it suffices to show that $\Pr\{\boldsymbol{Z} \mid \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\} = \Pr\{\boldsymbol{Z} \mid \boldsymbol{s}(\boldsymbol{x})\}$. From (1), $\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{d}, \boldsymbol{x}\right) = \Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right)$, so that

$$\begin{aligned}&\Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\right\} = \mathrm{E}\left\{\Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{d}, \boldsymbol{x}\right\} \,\middle|\, \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\right\} \\ &\quad= \mathrm{E}\left\{\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right) \,\middle|\, \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\right\} = \mathrm{E}\left\{e_g(\boldsymbol{x}) \,\middle|\, \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\right\}.\end{aligned}$$

As $e_g(\boldsymbol{x})$ is part of $\boldsymbol{s}(\boldsymbol{x}) = \left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$,

$$\begin{aligned}&\mathrm{E}\left\{e_g(\boldsymbol{x}) \,\middle|\, \boldsymbol{d}, \boldsymbol{s}(\boldsymbol{x})\right\} = e_g(\boldsymbol{x}) = \mathrm{E}\left\{e_g(\boldsymbol{x}) \,\middle|\, \boldsymbol{s}(\boldsymbol{x})\right\} \\ &= \mathrm{E}\left\{\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right) \,\middle|\, \boldsymbol{s}(\boldsymbol{x})\right\} = \Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{s}(\boldsymbol{x})\right\},\end{aligned}$$

proving (6). Then (7) follows from (2) because $\Pr\left\{Z_g = 1 \,\middle|\, \boldsymbol{s}(\boldsymbol{x})\right\} = \mathrm{E}\left\{\Pr\left(Z_g = 1 \,\middle|\, \boldsymbol{x}\right) \,\middle|\, \boldsymbol{s}(\boldsymbol{x})\right\}$. Finally, (8) follows from Proposition 1 with $\boldsymbol{x}$ replaced by $\left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$. □

3.3. Checking Covariate Balance with Aliasing

Blocking entails selecting individuals and grouping them into blocks that balance $\boldsymbol{x}$ and are homogeneous in certain features $\boldsymbol{f}(\boldsymbol{x})$ of $\boldsymbol{x}$. Let b denote the block to which an individual is assigned. Let $\mathcal{G}_b \subseteq \{1, \ldots, G\}$ be the groups present in block b. For instance, in Table 4, the 700 blocks b of type 2 all have $\mathcal{G}_b = \{1, 3, 5, 7\}$ in the first column of Table 3.

If blocks were homogeneous in $\left\{\boldsymbol{e}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x})\right\}$, then (5) would hold, and this motivates Definition 3. Recall that A is independent of B if and only if $\mathrm{E}\{c(A) \mid B\} = \mathrm{E}\{c(A)\}$ for all functions $c(\cdot)$. (More precisely: this condition must hold for almost all B and for all measurable $c(\cdot)$ for which the expectations exist, but we will not repeat this.)

Definition 3. The distribution of $\boldsymbol{x}$ is balanced within blocks if $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{Z} \mid b$, or equivalently if $\mathrm{E}\left\{c(\boldsymbol{x}) \mid b, Z_g = 1\right\} = \mathrm{E}\{c(\boldsymbol{x}) \mid b\}$ for every $c(\cdot)$ and b, or equivalently if for every $c(\cdot)$ and b,

$$\sum_{g \in \mathcal{G}_b} h_g \,\mathrm{E}\left\{c(\boldsymbol{x}) \mid b, Z_g = 1\right\} = 0 \text{ for every } h_g \text{ such that}$$

$$0 = \sum_{g \in \mathcal{G}_b} h_g. \tag{9}$$

A block design, with or without aliasing, must be checked for adequacy by checking whether $\boldsymbol{x}$ is balanced by blocking; see, for instance, Yu (2021) for a modern discussion and example.

The situation is different for eligibility covariates, $\boldsymbol{w}$, that are aliased with groups. Eligibility covariates $\boldsymbol{w}$ are always imbalanced, because they determine $\boldsymbol{Z}$; so, they never satisfy Definition 3. However, if $\beta(\boldsymbol{x}, \boldsymbol{w}) = \sum_{m=1}^{M} \lambda_m \delta_m(\boldsymbol{x}, \boldsymbol{w})$ for known $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ and unknown λ_m, as in Section 1, then it is natural to check $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ for balance with respect to a specific contrast h_g, $g = 1, \ldots, G$, say column (7) in Table 1. This motivates Definition 4, which is illustrated in Section 4.1.

Definition 4. A function $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ of unaliased covariates $\boldsymbol{x}$ and aliased covariates $\boldsymbol{w}$ is balanced with respect to contrast $h_g, g = 1, \ldots, G$, in block b if:

$$0 = \sum_{g \in \mathcal{G}_b} h_g \text{ and } \sum_{g \in \mathcal{G}_b} h_g \,\mathrm{E}\left\{\delta_m(\boldsymbol{x}, \boldsymbol{w}) \mid b, Z_g = 1\right\} = 0. \tag{10}$$

The important distinction between Definitions 3 and 4 is that (9) holds for *all* contrasts $h_g, g = 1, \ldots, G$ and *all* functions $c(\cdot)$, while (10) holds for a specific function $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ and a specific contrast $h_g, g = 1, \ldots, G$.

4. Benefits Duration and Weeks of Unemployment

4.1. Checking Covariate Balance

Figure 1 checks the covariate age, x_3, for balance in each of the six types of blocks and for $h_g = 1$ or $h_g = -1$ in Table 4. Here, covariate balance refers to a particular contrast h_g, as in Definition 4. A difference-in-differences effect estimate in a block type in Table 4 will compare individuals with $h_g = 1$ to individuals with $h_g = -1$, and in each block type in Figure 1, age does look balanced for $h_g = 1$ or $h_g = -1$. For an unaliased covariate like age in $\boldsymbol{x}$, balance could be checked over four groups defined by $g \in \mathcal{G}_b$ — that is balance could be checked for every contrast in Definition 3, rather than for the two groups defined by $h_g = 1$ or $h_g = -1$ in Table 4, but this is not possible for aliased covariates.

The function $\beta(\boldsymbol{x}, \boldsymbol{w}) = \sum_{m=1}^{M} \lambda_m \delta_m(\boldsymbol{x}, \boldsymbol{w})$ cannot be checked for balance when the $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ are known but the λ_m are unknown. We can check the $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ for balance with respect to a contrast, one $\delta_m(\boldsymbol{x}, \boldsymbol{w})$ at a time. This is similar to the common practice of checking observed covariates $\boldsymbol{x}$ for balance in a matched treatment/control comparison by checking individual covariates, x_m, for balance, and checking interactions, $x_m x_{m'}$, for treatment/control balance. Some terms do not need to be checked, because the design forces balance. The design forces other terms to be imbalanced, so that, miraculous cancellations aside, such a term would cause bias if present in $\beta(\boldsymbol{x}, \boldsymbol{w})$. Finally, it is an empirical matter whether the design has balanced certain other terms; typically, these are terms that involve both $\boldsymbol{x}$ and $\boldsymbol{w}$. As an example, consider blocks of type 2, which estimate the conditional main effect of B-versus-b at R, as represented by the contrast h_g, $g = 1, \ldots, G$ given by column (7) or $Bb@R$ in Table 1. In Table 1, contrast (3) for w' is orthogonal to contrast (7) or $Bb@R$ for the conditional main effect in blocks of type 2. Because it is balanced by design, of course we do not need to check a term in $\beta(\boldsymbol{x}, \boldsymbol{w})$ that depends only on w' for balance with respect to h_g in blocks of type 2. Similarly, the contrast (6) is orthogonal to the LE-by-Time interaction, $w' \times w'''$, or contrast (3) $\times$ (5), so a term in $\beta(\boldsymbol{x}, \boldsymbol{w})$ that depends only on (w', w''') does not need to be checked for balance. Also, columns (4) and (7) are orthogonal, so IU or w'' does not need to be checked for balance. Conversely, contrast (7) is not orthogonal to the interaction (4) $\times$ (5) or $w'' \times w'''$, so a term in $\beta(\boldsymbol{x}, \boldsymbol{w})$ that depends nontrivially on the value of the contrast (4) $\times$ (5) is necessarily imbalanced with respect to h_g in (7). In brief, the design itself makes some balance checks into a foregone conclusion.

The situation is different for the interaction covariate "age $\times$ LE" or $x_3 \times w'$. Like age itself, $x_3 \times w'$ may be balanced or not over $h_g = 1$ or $h_g = -1$, and we need to look at the data to judge whether matching has balanced $x_3 \times w'$ over $h_g = 1$ or $h_g = -1$. This covariate $x_3 \times w'$ is checked in Figure 2, consistent with Definition 4. The example discussed in the previous paragraph focused on blocks of type 2, but Figure 2 shows all six types of blocks. Because $w' = \pm 1$, both positive and negative values of $x_3 \times w'$ may appear in Figure 2, although only $w' = 1$ appears in block type 2 so only positive values appear Figure 2, and only $w' = -1$ appears in block type 5, so only negative values appear in Figure 2. In every block, however, individuals with positive and negative contrast values, $h_g = 1$ or $h_g = -1$, have similar distributions of $x_3 \times w'$; so, an imbalance in $x_3 \times w'$ cannot bias the contrast. Although $\beta(\boldsymbol{x}, \boldsymbol{w})$ is a weighted sum of many terms, a term of the form $x_3 \times w'$ seems likely to approximately cancel when the contrast $h_g = 1$ or $h_g = -1$ is applied, that is, when we compare individuals with $h_g = 1$ to individuals with $h_g = -1$ inside the same block.

Table 5 resembles Figure 1, but refers to many covariates. Table 5 compares the distribution of x_k for individuals with positive and negative contrast values, $h_g = 1$ or $h_g = -1$, to complete randomization of x_k to $h_g = 1$ or $h_g = -1$. This is done separately in each of the six block types, as the conditional

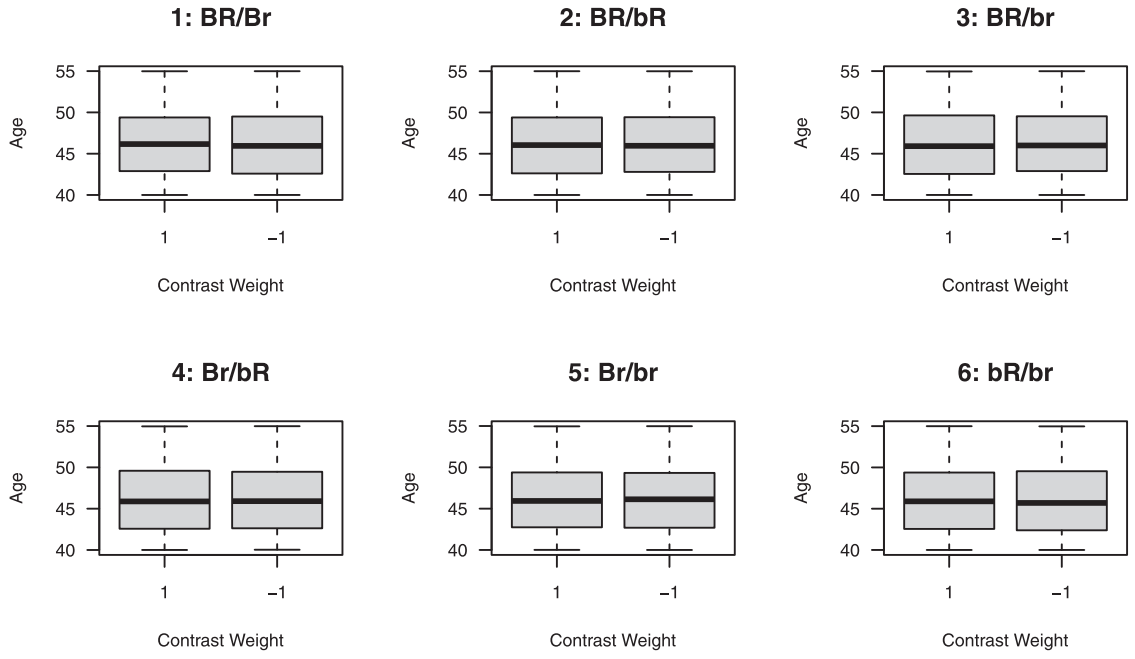

Figure 1. Balance of the covariate "age" by contrast value, +1 or −1, in each of six types of blocks.

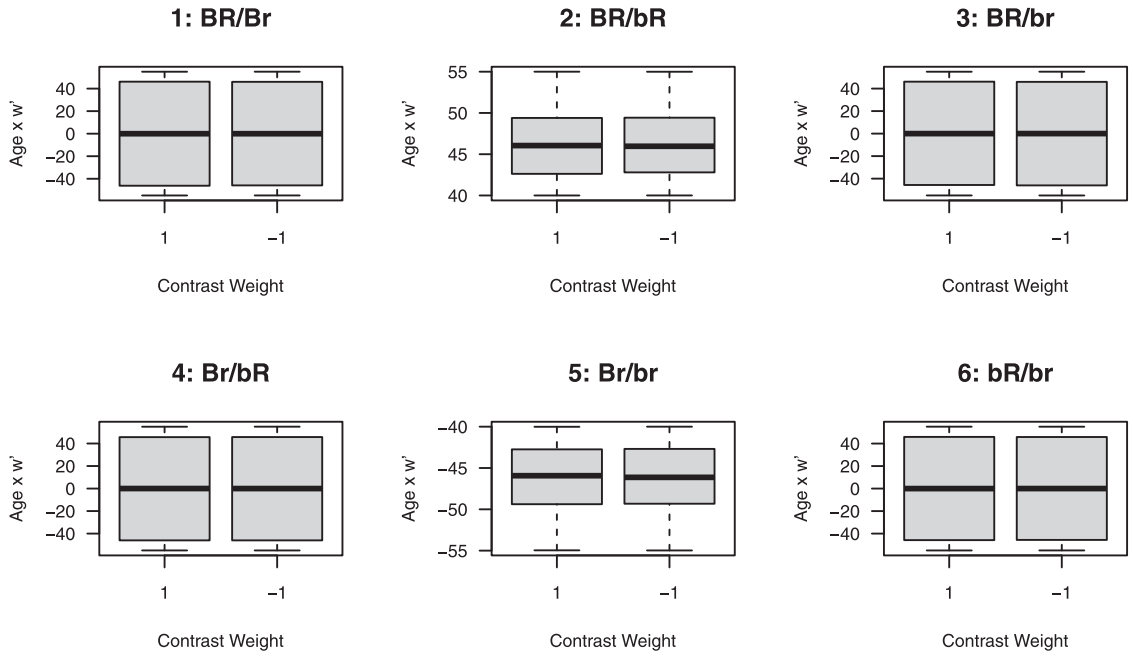

Figure 2. Balance of the covariate "age × LE" or $x_3 \times w'$ by contrast value, +1 or −1, in each of six types of blocks.

Table 5. p-values checking covariate imbalance, $h_g = 1$ versus $h_g = -1$, compared with complete randomization, in each of 6 block types, plus *combining* (C) across 6 block types using the truncated product method.

Covariate	6 types of blocks						
	1 BR/Br	2 BR/bR	3 BR/br	4 Br/bR	5 Br/br	6 bR/br	C
Temporary layoff	0.78	0.60	0.94	0.90	0.68	0.65	1.00
Female	0.94	1.00	0.82	1.00	0.97	0.88	1.00
Age	0.58	0.99	0.67	0.94	0.77	0.85	1.00
Secondary education	1.00	0.24	0.86	0.74	0.36	0.49	1.00
Tertiary education	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Apprenticeship	0.90	0.67	0.70	0.58	0.83	0.87	1.00
Married	0.81	0.97	0.94	0.90	0.97	0.94	1.00
Single	0.46	0.95	0.61	0.87	0.65	0.91	1.00
Divorced	1.00	0.92	0.67	0.79	0.75	0.79	1.00
Female x married	0.82	0.33	1.00	0.85	0.75	0.76	1.00
Female x single	0.93	0.42	1.00	1.00	0.93	0.92	1.00
Female x divorced	0.69	0.64	0.28	1.00	0.69	0.90	1.00
Blue collar job	0.63	0.79	0.97	0.54	0.90	0.57	1.00
Seasonal job	0.70	0.97	0.97	0.94	0.49	0.64	1.00
Manufacturing job	0.72	0.57	0.96	0.96	0.58	0.53	1.00
Prior wage	0.94	0.52	0.66	0.88	0.91	0.79	1.00
(LE) prior wage ≤ 12610	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Relative employment (RE)	0.90	0.98	0.85	0.91	0.93	0.95	1.00
(IU) RE ≥ 40%	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Worked 3 of 5 years	0.97	1.00	0.97	0.81	0.81	0.97	1.00
Age ≥ 50	0.71	0.93	0.96	0.71	0.89	0.68	1.00

mean effect contrasts are different in different block types. Having computed one p-value for each block type, the six within-type p-values are combined over block types using Zaykin et al.'s (2002) truncated product method of combining independent p-values, as implemented in the `truncatedP` method in the `R` package `sensitivitymv` with the default truncation of 0.2. This method defines a new statistic as the product of all p-values less than or equal to the truncation point, or 1 if no p-value is below the truncation point. The null distribution of this new statistic is determined, and that yields its p-value, which is reported in column C of Table 5. If the truncation point were 1 instead of 0.2, then the method would be equivalent to Fisher's method for combining independent p-values. As all of the p-values in Table 5 are above 0.2, the combined p-value is always 1. The balance of covariates in Table 5 is much better than expected from random assignment to $h_g = 1$ or $h_g = -1$.

Several aspects of Table 5 serve to illustrate Definition 4. In Definition 4, a covariate may be balanced with respect to contrast h_g over several groups even though some individual groups do not overlap with respect to that covariate. The covariate w' or LE is perfectly balanced because the same number of individuals in group $h_g = 1$ have $w' = 1$ as in group $h_g = -1$. The same is true of the covariate w'' or IU, and of the time period w'''. Both w' and w'' are two-level covariates formed by cutting continuous variables, namely "prior wage" for w' and "relative employment" for w''. Because eligibility cuts continuous covariates, the support of the continuous covariates does not overlap in some pairs of treatment groups. Nonetheless, these continuous covariates are balanced with respect to h_g in the sense of Definition 4: the contrast groups, $h_g = 1$ or $h_g = -1$, are balanced with respect to the continuous covariates by merging two groups with nonoverlapping support. An estimator of a treatment effect is a contrast in outcomes that will compare outcomes in the $h_g = 1$ or $h_g = -1$ groups, and that comparison is balanced with respect to covariates.

Table 6 checks the balance $x_k \times w'$, rather than x_k in Table 5. The balance is much better than is expected from complete randomization of individuals to $h_g = 1$ or $h_g = -1$; so, covariates of the form $x_k \times w'$ are not likely to bias an estimate that compares individuals with $h_g = 1$ and $h_g = -1$. Appendix Tables 2–5 are similar to Table 6, but explore other covariate interactions. Appendix Tables 2 and 3 concern interactions that *could* be balanced given the alias structure in Table 4, and all of those interactions are balanced. Appendix Tables 4 and 5 concern interactions that *cannot* be balanced given the alias structure in Table 4, and all of those interactions are severely

imbalanced. In the four tables of covariates or interactions that could be balanced, there are a total of 504 covariate/block-type situations; there, the median of 504 p-values is 0.8871 and the minimum is 0.2379, so the balance is much better than expected by complete randomization. In common practice, a linear model would not include so many covariates and interactions; yet here, all 504 are better balanced than by complete randomization. In the two tables where aliasing forces imbalances, there are 252 covariate/block-type interactions, with a median p-value of 0.0000 and an upper quartile of 0.5708. The upper quartile of 0.5708 reflects the fact that comparisons in certain block types are not aliased by eligibility covariates that create aliasing in other block types. For example, as in Section 2.2, low earnings or LE is exactly matched in each block of type 2, but it can never be exactly matched for all four individuals in blocks of type 3. Covariates and interactions that were not aliased were balanced by matching, but covariates that are aliased cannot be controlled by any statistical method.

Table 6. p-values checking covariate imbalance, $h_g = 1$ versus $h_g = -1$, for covariates interacted with LE, $x_k \times w'$, compared with complete randomization, in each of six block types, plus combining across six block types (C).

Covariate	6 types of blocks						
	1 BR/Br	2 BR/bR	3 BR/br	4 Br/bR	5 Br/br	6 bR/br	C
Temporary layoff	0.62	0.60	0.33	0.38	0.68	0.73	1.00
Female	0.94	1.00	0.76	1.00	0.97	0.88	1.00
Age	0.87	0.99	0.81	0.94	0.77	0.97	1.00
Secondary education	0.94	0.24	0.41	0.88	0.36	0.55	1.00
Tertiary education	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Apprenticeship	0.62	0.67	0.79	0.57	0.83	0.82	1.00
Married	0.82	0.97	0.94	0.91	0.97	0.94	1.00
Single	0.91	0.95	0.79	0.85	0.65	1.00	1.00
Divorced	0.71	0.92	0.88	0.85	0.75	0.85	1.00
Female x married	0.88	0.33	0.94	0.52	0.75	1.00	1.00
Female x single	0.97	0.42	0.71	1.00	0.93	0.97	1.00
Female x divorced	0.65	0.64	0.82	0.82	0.69	0.82	1.00
Blue collar job	0.73	0.79	0.97	0.60	0.90	0.68	1.00
Seasonal job	0.71	0.97	0.97	0.94	0.49	0.65	1.00
Manufacturing job	0.79	0.57	0.62	0.97	0.58	0.65	1.00
Prior wage	0.81	0.52	0.71	0.81	0.91	0.95	1.00
(LE) Prior wage $\leq$ 12610	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Relative employment (RE)	0.64	0.98	0.88	0.88	0.93	0.99	1.00
(IU) RE $\geq$ 40%	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Worked 3 of 5 years	0.97	1.00	0.97	0.82	0.81	0.97	1.00
Age $\geq$ 50	0.82	0.93	0.97	0.50	0.89	0.97	1.00

4.2. Examining the Outcome: Duration of Unemployment

To concisely illustrate data analyses, focus on blocks of types 2 and 5; however, the same methods apply to other block types. Figure 3 compares two direct estimates of the effect of the increase in benefits duration, B-versus-b, from blocks of types 2 and 5. However, block types 2 and 5 are different. In blocks of type 2, the replacement rate was increased in the after period, R, while in blocks of type 5 the replacement rate not increased in the after period; see columns (7) and (10) in Table 1. The two left panels of Figure 3 show the duration of unemployment, capped at 104 weeks or two years. The contrast weights, $h_g = 1$ or $h_g = -1$, appear above the boxplots. An increase in benefit duration B in the after period is associated with a somewhat longer duration of unemployment in the after period, whether the replacement rate is increased (type 2 blocks) or not (type 5 blocks). The distributions have long tails, somewhat obscuring the typical durations.

Each block produces one difference-in-differences, or 700 difference-in-differences from blocks of type 2 and 700 from blocks of type 5. Each of 700 difference-in-differences contrasts two individuals with $h_g = 1$ to two individuals with $h_g = -1$, asking whether duration of unemployment is different at $h_g = 1$ and $h_g = -1$. Combining blocks of types 2 and 5, the $1400 = 700 + 700$ differences-in-differences are plotted in the final boxplot. To see the typical difference clearly, the vertical axis has its tails symmetrically transformed beyond $\pm\beta$ using a $p = -1$ transformation, where β is the 0.8 quantile of the absolute difference-in-differences (Rosenbaum 2022). The median of the 1400 untransformed differences-in-differences is 2.3 weeks, the quartiles are -7.6 and 12.3, and the 10% and 90% points are -19.4 and 23.8.

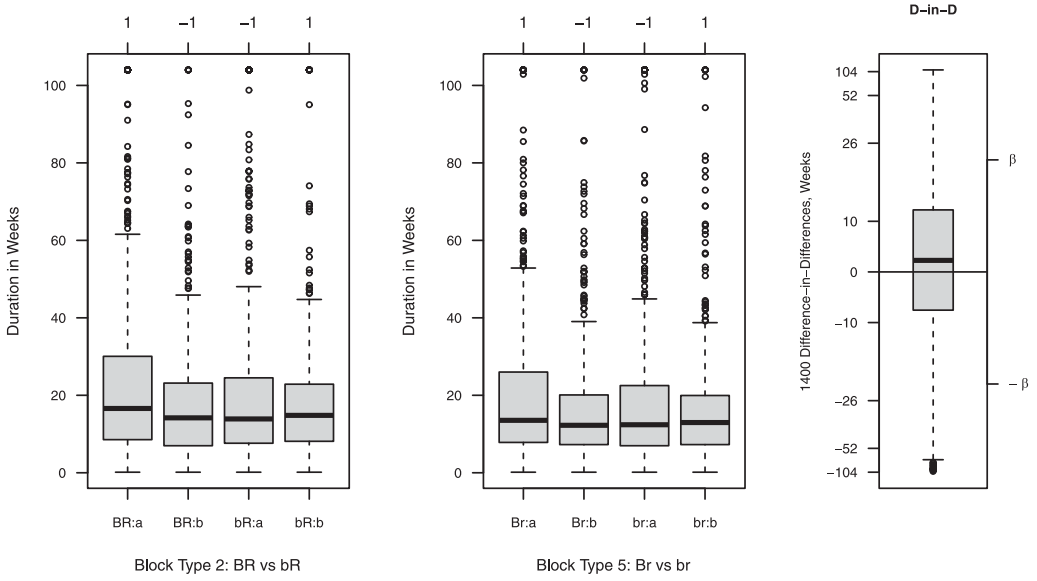

Figure 3. Duration of unemployment, with $h_g = 1$ or $h_g = -1$ above boxplots.

Blocks of type 2 estimate the effect of an increase in benefits duration, B-vs-b, at the high level, R, of the replacement rate, while blocks of type 5 refer to the low level, r, of the replacement rate. These two difference-in-differences for blocks of types 2 and 5 are estimating the same effect if there is no $Bb \times Rr$ interaction; otherwise, the shift in the distribution from blocks of type 2 to blocks of type 5 is the interaction. As seen in Table 4, these effects from blocks of types 2 and 5 are not aliased with the LE × IU interaction.

Each of the 1400 blocks of size 4 provides an estimate of one effect in columns (6) to (11) of Table 1 by comparing two people assigned to a group g with $h_g = 1$ to two people assigned to a group g with $h_g = -1$. Compared to a random allocation of 2 of 4 individuals to $h_g = 1$, the 1400 difference-in-differences are high, with one-sided p-value of 1.1×10^{-16}. The comparison is insensitive to a bias that increases the odds of assignment to $h_g = 1$ rather than $h_g = -1$ by a factor of $\Gamma = 1.6$, as the upper bound on the p-value is then 0.035. This calculation uses methods in Rosenbaum (2018, 2024) as implemented in the `R` function `gwgtRank` in the `weightedRank` package. In a matched pair, a bias of $\Gamma = 1.6$ is equivalent to an unobserved covariate that doubles the odds of treatment and increases the odds of a positive pair difference in outcome by 5-fold (Rosenbaum and Silber 2009). This sensitivity analysis presumes that (6) and (7) may be false, but would have been true if the conditioning had included also the unobserved covariate. In other words, blocks of types 2 and 5 provide some evidence of a main effect of increased benefits duration on the duration of unemployment, an effect that is clearly not due to chance, and is not easily explained by small unmeasured biases in group assignment.

The differences-in-differences are slightly but not significantly longer in the 700 blocks of type 2, with an increase in the replacement rate, than in the 700 blocks of type 5, without an increase: Wilcoxon's rank sum test yields a p-value of 0.13 with a Hodge-Lehmann estimated difference of 1.28 weeks and a 95% confidence interval of $[-0.43, 3.06]$ weeks. In other words, there is little or no indication of a $Bb \times Rr$ interaction in blocks of types 2 and 5.

5. Construction of Balanced Blocks of Maximal Size

The block design in Table 4 was assembled in three steps. Steps 1 and 2 use an extension of cardinality matching (Zubizarreta, Paredes, and Rosenbaum 2014) to select units; then, Step 3 uses minimum distance matching to put the selected units into blocks so that the blocks are as homogeneous as possible with respect to covariates. Steps 1 and 2 constitute a p-way template-balanced partitioning program, where each treatment group is partitioned into p samples of maximal size balanced with respect to a template. As seen in Figures 1–2, Table 5 and in other tables in the appendix, the observed covariates exhibit good balance, indeed better balance than is expected from complete randomization of these same individuals to these same groups. Of course, as always, balance for observed covariates does not imply balance for covariates that were not observed.

In forming a single matched sample, cardinality matching picks the largest control group that satisfies certain requirements for covariate balance; see Niknam and Zubizarreta (2022) for a practical exposition. Cardinality matching differs from minimum distance matching, in which the size of the design is fixed in advance and a minimum distance match of that size is constructed.

Recent applications pick controls to resemble the distribution of covariates in a target or "template" population; see Silber et al. (2014) and Cohn and Zubizarreta (2022). This permits more complex designs, because the size of the template no longer plays a role in matching.

In cardinality matching, a balance constraint is a linear inequality constraint. Let s be the size of the matched control group. Two inequality constraints can express a constraint on an absolute value: specify an $\epsilon > 0$; then two inequalities, $A \leq s\epsilon$ and $-A \leq s\epsilon$, jointly require $|A| \leq s\epsilon$. For example, we might require the mean age among selected controls to differ from the mean age in the template by at most one year, $\epsilon = 1$, and that requirement A can be expressed as a requirement on a linear quantity, $|A| \leq s\epsilon$, where $A = sB - \sum_{i=1}^{I} c_i a_i$, B is the mean age in the template, a_i is the age of the ith of I available controls, $c_i = 1$ if control i is selected into the control group and $c_i = 0$ otherwise, and $s = \sum_{i=1}^{I} c_i$. There are many such linear requirements for many covariates, all with the same c_i and s. Cardinality matching picks the $c_i \in \{0, 1\}$ to maximize $s = \sum_{i=1}^{I} c_i$ subject to these linear constraints.

Building the design in Table 4 requires an adjustment to cardinality matching. There are $4 \times 2 = 8$ groups: BR, Br, bR, br, $\overline{\text{BR}}$, $\overline{\text{Br}}$, $\overline{\text{bR}}$, and $\overline{\text{br}}$. Focus on one group, say BR. Group BR needs to be represented three times in three separate block types of size s, one for blocks of type 1, one for blocks of type 2, and one for blocks of type 3. The situation is essentially the same for the other seven groups. Focus on group BR which contains I individuals before matching. Let $c_{ip} = 1$ if individual i from BR is selected for group p, $p = 1, \ldots, p$, where $P = 3$. To ensure that a member of BR is selected for at most one group, there is a linear constraint $c_{i1} + \cdots + c_{iP} \leq 1$ for $i = 1, \ldots, I$. For the covariate age, there are now $P = 3$ constraints instead of one, $\left|A_p\right| \leq s\epsilon$ with $A_p = sB - \sum_{i=1}^{I} c_{ip} a_i$ for $p = 1, \ldots, P$. Finally, the P groups must each be of size s, so there are p size constraints, $s = \sum_{i=1}^{I} c_{ip}$ for $p = 1, \ldots, P$. The problem in step 1 is to pick $c_{ip} \in \{0, 1\}$ to maximize s subject to the stated constraints. That produces $P = 3$ groups from BR that each resemble the template in terms of the distribution of covariates.

The problem just described is solved $4 \times 2 = 8$ times, for BR, Br, bR, br, $\overline{\text{BR}}$, $\overline{\text{Br}}$, $\overline{\text{bR}}$, and $\overline{\text{br}}$. This yields eight values of s, and the minimum value, $\bar{s}$, of s is selected. This completes Step 1.

Step 2 builds $4 \times 2 = 8$ balanced groups all of the same size, $\bar{s}$. In Step 2, for each of the $4 \times 2 = 8$ groups, the problem in Step 1 is solved again, but now with the size $\bar{s}$ that was determined in Step 1; that is, with the fixed constraint $\bar{s} = \sum_{i=1}^{I} c_{ip}$ for $p = 1, \ldots, P$. This can be done in several ways that differ slightly and are described in the Appendix. For example, with $\bar{s}$ fixed, one could optimize another quantity, such as the sum of suitably scaled ϵ's over all of the covariates. See the appendix for options and details.

At the end of Step 2, there are 24 balanced groups. Four balanced groups are assigned to block types 1–6 in Table 4. Step

3 assembles the $4\bar{s}$ individuals in each block type into $4\bar{s}$ blocks of 4 individuals, one from each of the four balanced groups in that block type. Within block type 2, $\overline{\text{BR}}$ is optimally paired with BR, to minimize a rank-based Mahalanobis covariate distance (Rosenbaum 2020, sec. 9.3). Importantly, because the eligibility variables—that is, $w' = \text{LE}$ and $w'' = \text{IU}$ and their continuous counterparts Prior Wage and Relative Employment in Table 5—are *not* aliased for $\overline{\text{BR}}$ and BR, the before-after pairing of BR to produce $\overline{\text{BR}} - \text{BR}$ pairs *can* pair for these covariates. Similarly, within block type 2, $\overline{\text{bR}}$ is optimally paired with bR. Then the $\overline{\text{BR}} - \text{BR}$ pairs are paired with the $\overline{\text{bR}} - \text{bR}$ pairs. In pairing pairs, the distance between two pairs is the sum of the four covariate distances that cross the pairs (see, Nattino et al. 2021). In pairing pairs, some eligibility covariates *cannot* be included, depending upon the block type.

6. Discussion: Recap and Extensions

6.1. Preference For Reduced Aliasing Among Interactions

In his discussion of the planning of observational studies, Cochran (1965, p. 236) wrote:

> …to a large extent, workers in observational research have tried to copy devices that have proved effective in controlled experiments … Dorn (1953) recommended that the planner of an observational study always ask himself the question, "How would the study be conducted if it were possible to do it by controlled experimentation?"

In designing a fractional factorial experiment, an investigator prefers a design in which fewer main effects and low-order interactions are aliased with each other, but instead are aliased with high-order interactions thought to be negligible. This principle is also relevant in observational studies in which eligibility for certain treatments is a deterministic function of certain eligibility covariates, so that the propensity score given all observed covariates is always 0 or 1.

In certain respects, the alias structure in Table 4 for conditional effects is preferable to the alias structure in Table 2. Policy makers may find Wu's conditional effects to be closer to their thinking than: (i) a main effect that averages effects over several policies, and (ii) interactions that ask whether splitting that average apart would reveal different effects among the policies just averaged. For example, $Bb@R$ in column (2) of Table 2 asks a question about whether there is an effect of an increase in benefits duration among individuals with low earnings (LE) who received an increase in the benefit amount (R). As is natural, that question is answered using only individuals with low earnings (LE). In other words, that comparison is exactly matched for LE in blocks of type 2, even though LE cannot be exactly matched when comparing all eight groups in Table 2. Moreover, in Table 4, $Bb@R$ is orthogonal to most, but not all, two-factor interactions. Each of the six columns of Tables 3 and 4 refers to an easily interpreted effect, each with different aliasing. For instance, in column 3 of Table 4, the rather natural comparison of increasing both benefits versus increasing neither benefit is not aliased with the LE×IU interaction.

6.2. Aspects Outside Classical Experimental Design

Certain aspects of our blocked design in Tables 3 and 4 appear to have no analogous form in experimental design, essentially because the observational treatment assignment is found by the investigator, not created by the investigator. An investigator would not create some of the structures that she may find. In particular, as noted in Section 5, our before-versus-after pairing of individuals in the same $w' \times w''$ category can and did match for both binary and continuous aliased covariates (binary LE and IU, and continuous Prior Wage and Relative Employment) that lack common support as $w' \times w''$ varies. One might say that in this difference-in-differences, the inter-temporal difference is more closely matched than the inter-eligibility difference.

6.3. Observational Studies with More Complex Aliasing

Is complex aliasing rare in observational studies? Is Lalive, Van Ours, and Zweimüller's (2006) study unusual in having complex aliasing? We suspect that complex aliasing is often unnoticed, but not unusual. We most readily recognize in data those patterns that we have already seen in mathematical structures. For instance, stronger resistance to aliasing is possible in Lalive, Van Ours, and Zweimüller's (2006) study by distinguishing its four years as four years, rather than two years before and two years after the policy change; see online Appendix 3. With four distinct years and treatment beginning at the beginning of year three, a temporal discontinuity is expected between year two and three, and that can be distinguished from differing linear time trends in eligible and ineligible groups.

Supplementary Materials

The online supplementary materials include appendices with additional information on the benefits program, balance checks, orthogonal contrasts, the three-step matching method, and the template. They also provide R code to replicate the analyses presented in the paper. Further details about the code, along with instructions for requesting the data, are also provided in the supplementary materials in the ACC form associated with this paper.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The authors gratefully acknowledge funding from the Patient Centered Outcomes Research Initiative (PCORI, award ME-2022C1-25648).

References

Bailey, R. A. (2008), *Design of Comparative Experiments*, Cambridge: Cambridge University Press. [4]

Cahuc, P., Carcillo, S., and Zylberberg, A. (2014), *Labor Economics*, Cambridge, MA: MIT Press. [3]

Card, D., and Krueger, A. B. (1994), "Minimum Wages and Employment," *The American Economic Review*, 84, 772–793. [1,5]

Cochran, W. G. (1965), "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society A*, 128, 234–266. [11]

Cohn, E. R., and Zubizarreta, J. R. (2022), "Profile Matching for the Generalization and Personalization of Causal Inferences," *Epidemiology*, 33, 678–688. [10]

Cook, T. D., Campbell, D. T., and Shadish, W. (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin. [2]

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory," (with Discussion), *Journal of the Royal Statistical Society*, Series B, 41, 1–15. [5,6]

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, 854–866. [6]

Lalive, R., Van Ours, J., and Zweimüller, J. (2006), "How Changes in Financial Incentives Affect the Duration of Unemployment," *Review of Economic Studies*, 73, 1009–1038. [2,4,11]

Meyer, B. D. (1995), "Natural and Quasi-Experiments in Economics," *Journal of Business & Economic Statistics*, 13, 151–161. [2]

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, 868–881. [5]

Nattino, G., Lu, B., Shi, J., Lemeshow, S., and Xiang, H. (2021), "Triplet Matching for Estimating Causal Effects with Three Treatment Arms," *Journal of the American Statistical Association*, 116, 44–53. [11]

Niknam, B. A., and Zubizarreta, J. R. (2022), "Using Cardinality Matching to Design Balanced and Representative Samples for Observational Studies," *Journal of the American Medical Association*, 327, 173–174. [10]

Rosenbaum, P. R. (2005), "Heterogeneity and Causality," *American Statistician*, 59, 147–152. [4]

——— (2018), "Sensitivity Analysis for Stratified Comparisons in an Observational Study of the Effect of Smoking on Homocysteine," *Annals of Applied Statistics*, 12, 2312–34. [10]

——— (2020), *Design of Observational Studies* (2nd ed.), New York: Springer. [11]

——— (2021), *Replication and Evidence Factors in Observational Studies*, Boca Raton, FL: CRC Press. [2]

——— (2022), "A New Transformation of Treated-Control Matched-Pair Differences for Graphical Display," *American Statistician*, 76, 346–352. [9]

——— (2024), "Bahadur Efficiency of Observational Block Designs," *Journal of the American Statistical Association*, 119, 1871–1881. [10]

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. [1,5,6]

——— (2023), "Propensity Scores in the Design of Observational Studies for Causal Effects," *Biometrika*, 110, 1–13. [1,6]

Rosenbaum, P. R., and Silber, J. H. (2009), "Amplification of Sensitivity Analysis in Matched Observational Studies," *Journal of the American Statistical Association*, 104, 1398–1405. [10]

Shaikh, A. M., and Toulis, P. (2021), "Randomization Tests in Observational Studies with Staggered Adoption of Treatment," *Journal of the American Statistical Association*, 116, 1835–1848. [2]

Silber, J. H., Rosenbaum, P. R., Ross, R. N., Ludwig, J. M., Wang, W., Niknam, B. A., et al. (2014), "Template Matching for Auditing Hospital Cost and Quality," *Health Services Research*, 49, 1446–1474. [10]

Stigler, G. J. (1946), "The Economics of Minimum Wage Legislation," *The American Economic Review*, 36, 358–365. [1]

Su, H., and Wu, C. J. (2017), "CME Analysis: A New Method for Unraveling Aliased Effects in Two-Level Fractional Factorial Experiments," *Journal of Quality Technology*, 49, 1–10. [2,3]

Wu, C. J. (2015), "Post-Fisherian Experimentation," *Journal of the American Statistical Association*, 110, 612–620. [2,3]

Wu, C. J., and Hamada, M. S. (2021), *Experiments* (3rd ed.), New York: Wiley. [2,3,4]

Ye, T., Keele, L., Hasegawa, R., and Small, D. S. (2024), "A Negative Correlation Strategy for Bracketing in Difference-in-Differences," *JASA*, 119, 2256–2268. [2]

Yu, R. (2021), "Evaluating and Improving a Matched Comparison of Antidepressants and Bone Density," *Biometrics*, 77, 1276–1288. [6]

Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H., and Weir, B. S. (2002), "Truncated Product Method for Combining p-values," *Genetic Epidemiology*, 22, 170–185. [8]

Zubizarreta, J. R., Paredes, R. D., and Rosenbaum, P. R. (2014), "Matching for Balance, Pairing for Heterogeneity in an Observational Study of the Effectiveness of For-Profit and Not-for-Profit High Schools in Chile," *Annals of Applied Statistics*, 8, 204–231. [4,10]