Optimal Pair Matching With Two Control Groups

Bo Lu and Paul R. Rosenbaum

In an effort to detect hidden biases due to failure to control for an unobserved covariate, some observational or nonrandomized studies include two control groups selected to systematically vary the unobserved covariate. Comparisons of the treated group and two control groups must, of course, control for imbalances in observed covariates. Using the three groups, we form pairs optimally matched for observed covariates—that is, we optimally construct from observational data an incomplete block design. The incomplete block design may use all available data, or it may use data selectively to produce a balanced incomplete block design, or it may be the basis for constructing a matched sample when expensive outcome information is to be collected only for sampled individuals. The problem of optimal pair matching with two control groups is shown by a series of transformations to be equivalent to a particular form of optimal nonbipartite matching, a problem for which polynomial time algorithms exist. In our examples, we implement the procedure using a nonbipartite matching algorithm due to Derigs. We illustrate the method with data from an observational study of the employment effects of the minimum wage.

Key Words: Assignment algorithm; Balanced incomplete block design; Bipartite matching; Multiple control groups; Nonbipartite matching; Observational studies.

1. INTRODUCTION

For observational studies of treatment effects that compare a treated group to two distinct control groups, an algorithm for optimal pair matching is developed in this article. Before developing the algorithm, Section 1 reviews relevant issues and presents a motivating example. The reasons for using two control groups in an observational study are briefly reviewed in Section 1.1. The role of matching in observational studies is briefly reviewed in Section 1.2. The method will be illustrated using Card and Krueger's (1994) study of the effects of the minimum wage on employment, and this study is discussed in Section 1.3.

Bo Lu is Assistant Professor, Center for Statistical Sciences, Brown University, Providence, RI 02912 (E-mail: bolu@stat.brown.edu). Paul R. Rosenbaum is Robert G. Putzel Professor, Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6302 (E-mail: rosenbaum@stat.wharton.upenn.edu).

Journal of Computational and Graphical Statistics, Volume 13, Number 2, Pages 422-434
DOI: 10.1198/1061860043470

1.1 REVIEW OF THE MAIN ISSUES: ROLE OF A SECOND CONTROL GROUP

In an observational study, subjects are not assigned to treatment or control at random, and may not be comparable prior to treatment, so that differing outcomes may not reflect effects caused by the treatment. Differences in observed pretreatment covariates are controlled by adjustments such as matching (Cochran 1965), but the possibility remains that the groups differ in terms of unobserved covariates which cannot be controlled by adjustments. Although most observational or nonrandomized studies of treatment effects use a single control group, some studies use two control groups in an effort to detect hidden biases from an unobserved covariate (Campbell 1969; Rosenbaum 1987, 2002a, sec. 8; Meyer 1995; Shadish, Cook, and Campbell 2002).

Campbell argued that one should select two control groups to systematically vary an unobserved covariate—that is, to select groups known to differ substantially on the covariate, even though the individual values of the covariate are unknown. His point was that if bias due to the unobserved covariate is responsible for the differing outcomes in treated and control groups, then this should be apparent, because the control groups should differ from each other. Though sometimes costly or difficult, the use of a second control group, if carefully selected, may render less plausible concerns about biases from certain unobserved covariates, thereby strengthening the evidence that the treatment caused its ostensible effects.

Among the many observational studies that have used two control groups are Seltser and Sartwell (1965); Weston and Mansinghka (1971); Roghmann and Sodeur (1972); Zabin, Hirsch, and Emerson (1989); Chang, Hwang, Wang, and Wang (1997); Wells et al. (1997); and Douglas and Carney (1998). It is uncommon for an observational study to have more than two control groups, although this is possible in principle.

1.2 REVIEW OF THE MAIN ISSUES: MATCHING IN OBSERVATIONAL STUDIES

Of course, when using two control groups, one must still adjust for observed covariates. Matching on observed covariates is a common method of adjustment, and it is often combined with further analytical adjustments. Rubin (1973, 1979) showed that regression analysis of matched samples is more robust to model misspecification than regression adjustment of random samples. Indeed, in some of his simulations, covariance adjustment unaided by matching increased the bias, whereas covariance adjustment of matched samples consistently reduced bias. Smith (1997) argued that if many potential controls are available, then using only comparable controls and discarding potential controls entirely unlike treated subjects can result in better control of bias with only slight loss of efficiency. Dehejia and Wahba (1999) illustrated the hazards of model-based adjustment unaided by matching by comparing experimental results with observational comparisons. The biases addressed by matching do not diminish with increasing sample size—they are $O(1)$—whereas the variance of an estimator does decrease—it is $O\left(\frac{1}{n}\right)$—so that as the sample size $n \to \infty$, the mean square error is dominated by the squared bias. Matching is also often used in research design to select a sample of individuals for whom costly additional information will be obtained; see Reinisch, Sanders, Mortensen, and Rubin (1995) and Silber et al. (2001) for two examples. Matching facilitates coordination of quantitative and qualitative research (Rosenbaum and Silber 2001).

The use of network flow optimization to form optimally matched samples was proposed by Rosenbaum (1989), and Bergstralh, Kosanke, and Jacobsen (1996) provide an implementation in SAS. Optimal matching has been used by Rephann and Isserman (1994), Nuako et al. (1998), Ghavamian et al. (1999), and Warner et al. (1999) and Silber et al. (2001).

1.3 MOTIVATING EXAMPLE: EFFECTS OF INCREASING THE MINIMUM WAGE

Economic theory predicts that minimum wage laws depress employment among low wage workers—that is, the theory predicts that the laws injure the people they are intended to benefit. In an effort to see whether this is so, Card and Krueger (1994) compared changes in employment at fast food restaurants in New Jersey, where the minimum wage had been increased on April 1, 1992, by nearly 20% from \$4.25 per hour to \$5.05 per hour, to similar restaurants in adjacent eastern Pennsylvania, where the minimum wage remained at \$4.25. They found no sign of depressed employment at New Jersey restaurants following the wage increase.

Although Burger Kings in New Jersey are quite similar to Burger Kings in Pennsylvania, and the economies of these two states are not sharply divided, it is nonetheless true that New Jersey and Pennsylvania differ in ways other than the minimum wage. For example, income and sales taxes differ in the two states. To gain insight into possible hidden biases, in one of several analyses, Card and Krueger (1994, tab. 3) compared restaurants affected by the minimum wage increase to two control groups not affected by the wage increase. One control group came from Pennsylvania, the other from unaffected restaurants in New Jersey. If an uncontrolled difference between Pennsylvania and New Jersey was masking the effects of the minimum wage increase that are predicted by economic theory, then Card and Krueger argued this should become apparent in the comparison with two control groups, because one of the control groups is in New Jersey. This is, again, Campell's notion of control groups selected to systematically vary an unobserved, uncontrolled covariate.

In New Jersey, some restaurants were paying a starting wage $\geq$ \$5.00 per hour before the increase in the minimum wage, so the new law did not require these restaurants to make substantial increases in their wages, whereas many other New Jersey restaurants had the old minimum \$4.25 as their starting wage. The three groups Card and Krueger compared were: New Jersey restaurants with a starting wage of \$4.25, New Jersey restaurants with a starting wage $\geq$ \$5.00 per hour, and Pennsylvania restaurants. In principle, if increasing New Jersey's minimum wage from \$4.25 to \$5.05 depressed employment, then the first group of restaurants should be affected but the other two groups should be largely unaffected. On the other hand, if restaurants in Pennsylvania are an inappropriate control group for restaurants in New Jersey, then we might reasonably hope to detect this by comparing Pennsylvania and New Jersey restaurants that were unaffected by the wage increase. In fact, the employment changes were not very different in these three groups of restaurants, so restaurants that were forced by the law to increase their starting wages by nearly 20% did not appear to experience a dramatic decline in employment when compared to two different control groups, and the control groups did not differ markedly from each other. By separating general differences between the states from the specific effects of the minimum wage, Card and Krueger strengthened their evidence that the minimum wage increase had little or no effect on employment.

Card and Krueger compared the three groups of restaurants without matching on covariates, although the restaurants differ in some ways. In particular, starting wages prior to the change in law were, by definition of the groups, starkly different. As an illustration of the matching technique, we will perform a matched analysis with three groups using Card and Krueger's data, an analysis whose conclusions turn out to be generally consistent with their unmatched comparison.

The comparison of three groups cannot reasonably be done as matched triples; rather, it must be done as an incomplete block design with matched pairs. The reason is that one wants to compare New Jersey restaurants which had been paying roughly the old minimum wage to similar Pennsylvania restaurants also paying roughly minimum wage; New Jersey restaurants paying the minimum to New Jersey restaurants paying much more; and New Jersey restaurants paying much more to Pennsylvania restaurants paying much more—these three comparisons cannot fit in a matched triple, but are easily accommodated in disjoint pairs. Moreover, pairs can be more closely matched for other covariates: incomplete block designs impose fewer constraints than complete block designs would on who can be matched to whom.

2. MATCHED PAIRS WITH THREE GROUPS

2.1 NOTATION

Three disjoint sets of units are given, $\mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$, $\mathcal{B} = \{\beta_1, \ldots, \beta_M\}$, $\mathcal{C} = \{\gamma_1, \ldots, \gamma_N\}$, with $N \geq M \geq K$, and a distance, $\delta(\cdot)$, defined for each pair—for instance, (α_i, β_j)—of units from different sets, for example, $\delta\{(\alpha_i, \beta_j)\}$. Notice that $\delta(\cdot)$ is a function of the pair (α_i, β_j). The distance is strictly positive, may be infinite, and need not satisfy the triangle inequality, so it need not be a distance in the topological sense. (Because the distance function $\delta(\cdot)$ has a finite domain—because it maps a finite set of possible pairs into the real numbers—it follows that if $\delta(\cdot)$ were not strictly positive, then it could be made strictly positive by adding a constant to $\delta(\cdot)$, and this would not change the matching that minimizes the total distance.)

Some α's will be paired with β's, and other α's will be paired with γ's, and some β's will be paired with γ's, but no unit will be used in more than one pair. This is the simplest form of incomplete block design—three groups in blocks of size two. The problem is to pick the best possible pairing, that is, to form the pairing to minimize the total distance within pairs.

Notice that this task is more complex than pairing two groups. With two groups, say $\mathcal{A}$

and $\mathcal{B}$, the problem would be which α's should be assigned to which β's; traditionally, this is called the optimal assignment problem or the optimal bipartite matching problem, and it is a standard problem. See Bertsekas (1991), or Ahuja, Magnanti, and Orlin (1993) for discussion of optimal assignment algorithms and see Rosenbaum (1989) and Bergstralh, Kosanke, and Jacobsen (1996) for discussion of their application to observational studies. With three groups, when pairing α_i, one has an additional choice, namely whether to pair using $\mathcal{B}$ or $\mathcal{C}$, and then to pick a particular $\beta \in \mathcal{B}$ or $\gamma \in \mathcal{C}$.

Suppose that we want $P_{\alpha\beta}$ pairs of the form $\left(\alpha_i, \beta_j\right)$, $P_{\alpha\gamma}$ pairs of the form $\left(\alpha_i, \gamma_j\right)$, and $P_{\beta\gamma}$ pairs of the form $\left(\beta_i, \gamma_j\right)$. Write $\mathcal{P}$ for a tripartite matching; that is, $\mathcal{P}$ is a set of $P_{\alpha\beta} + P_{\alpha\gamma} + P_{\beta\gamma}$ disjoint pairs with the required number of each type, so each element $\pi \in \mathcal{P}$ is one pair, and there are $P_{\alpha\beta}$ pairs $\pi \in \mathcal{P}$ of the form $\pi = \left(\alpha_i, \beta_j\right)$, and so on. Call $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$ the *size* of the matching $\mathcal{P}$. This size is fixed: all matchings we consider are of size $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$. The distance for this pairing is $\Delta(\mathcal{P}) = \sum_{\pi \in \mathcal{P}} \delta(\pi)$. A tripartite matching is *optimal* if it minimizes the distance $\Delta(\mathcal{P})$ over all $\mathcal{P}$ with the required size, $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$. The optimal matching problem is *feasible* if there exists such a matching of the required size with finite distance, $\Delta(\mathcal{P}) < \infty$; otherwise, it is infeasible.

Often, we want a balanced incomplete block design with $P_{\alpha\beta} = P_{\alpha\gamma} = P_{\beta\gamma}$, but this is not essential, nor is it always what we want. Write $k = P_{\alpha\beta} + P_{\alpha\gamma}$, $m = P_{\alpha\beta} + P_{\beta\gamma}$, and $n = P_{\alpha\gamma} + P_{\beta\gamma}$. Obviously, the triple $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$ is possible only if $K \geq k$, $M \geq m$, and $N \geq n$, and we assume these inequalities are satisfied.

2.2 MATCHING WITHOUT GROUPS: NONBIPARTITE MATCHING

Our strategy is to transform the tripartite matching problem into a different problem for which good algorithms exist. Suppose there is one set $\mathcal{D}$ containing an even number of elements, with a distance $\delta\{\cdot\}$ defined on all pairs $\left(\theta_i, \theta_j\right)$ with $\theta_i, \theta_j \in \mathcal{D}$. The so-called nonbipartite matching problem is to divide $\mathcal{D}$ into disjoint pairs in such a way as to minimize the total distance within pairs. It turns out that this problem is somewhat more complex than the bipartite matching problem, but good algorithms do exist to solve it. If Q units are to be matched in pairs, an optimal weighted matching can be found in $O\left(Q^3\right)$ operations; see Papadimitriou and Steiglitz (1982, p. 266). For comparison, multiplying two $Q \times Q$ matrices in the conventional way also uses $O\left(Q^3\right)$ operations. Derigs (1988) developed one good algorithm together with published FORTRAN code. Lu, Zanutto, Hornik, and Rosenbaum (2001) used Derigs' algorithm to quickly form 260 pairs from 521 subjects. Snyder and Steele (1990) showed that optimal matching is much better than greedy matching that forms pairs one at a time.

Our strategy is: (1) to associate the tripartite matching problem, with three groups $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, with a particular nonbipartite matching problem with a single group $\mathcal{D}$; (2) to prove that they have the same optimal solutions; and (3) to apply Derigs' (1988) algorithm. It is easiest to do this in two steps, first for a simple, special case in Section 2.3, and then by transforming the general case into the special case in Section 2.4.

2.3 A SIMPLE CASE: EQUAL GROUPS, BALANCED DESIGN

The key idea is clearest in a special case, so this case is presented first, and generalized in Section 2.4. In the special case, the groups are initially the same size, $K = M = N$ with N even, every unit is matched, so $k = K$, $m = M$, and $n = N$, and the resulting design is a balanced incomplete blocks design, so $P_{\alpha\beta} = P_{\alpha\gamma} = P_{\beta\gamma} = \frac{N}{2}$. Let $\mathcal{D} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C} = \{\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_N, \gamma_1, \ldots, \gamma_N\}$. Now $\delta(\cdot)$ is defined for some pairs of elements from $\mathcal{D}$ but not for other pairs. Extend the definition of $\delta(\cdot)$ to all pairs of elements from $\mathcal{D}$ by setting $\delta\{(\alpha_i, \alpha_j)\} = \delta\{(\beta_i, \beta_j)\} = \delta\{(\gamma_i, \gamma_j)\} = \infty$.

Claim 1. *Let $\mathcal{P}$ be an optimal nonbipartite matching in $\mathcal{D}$. If $\Delta(\mathcal{P}) < \infty$, then $\mathcal{P}$ is an optimal tripartite matching; if $\Delta(\mathcal{P}) = \infty$, then there is no feasible tripartite matching.*

Proof: Every tripartite matching is also a nonbipartite matching from $\mathcal{D}$. Conversely, let $\mathcal{P}^*$ be any nonbipartite matching of $\mathcal{D}$ and write $p_{\zeta\omega}$ for the number of pairs of type (ζ_i, ω_j) in $\mathcal{P}^*$; for example, $p_{\alpha\beta}$for the number of pairs of type (α_i, β_j). Does $p_{\alpha\beta} = P_{\alpha\beta} = \frac{N}{2}$, and so on as required? If $\Delta(\mathcal{P}^*) < \infty$, then $\mathcal{P}^*$ contains no pairs of the form (α_i, α_j), (β_i, β_j), or (γ_i, γ_j), and $p_{\alpha\alpha} = p_{\beta\beta} = p_{\gamma\gamma} = 0$. Moreover, because every unit in $\mathcal{D}$ is matched in $\mathcal{P}^*$, it follows that $N = p_{\alpha\beta} + p_{\alpha\gamma}$, $N = p_{\alpha\beta} + p_{\beta\gamma}$, and $N = p_{\alpha\gamma} + p_{\beta\gamma}$, so that $p_{\alpha\beta} = p_{\alpha\gamma} = p_{\beta\gamma} = \frac{N}{2}$, as required, and $\mathcal{P}^*$ is indeed a tripartite matching of the required size $(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}) = (\frac{N}{2}, \frac{N}{2}, \frac{N}{2})$. This proves that an optimal nonbipartite matching $\mathcal{P}$ from $\mathcal{D}$ with $\Delta(\mathcal{P}) < \infty$ is also an optimal, feasible tripartite matching, whereas if $\Delta(\mathcal{P}) = \infty$ there is no tripartite matching with finite total distance. □

2.4 THE GENERAL CASE

In the general case, it is not necessary to match all the units, $K \geq k$, $M \geq m$, and $N \geq n$, the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$, need not have the same number of units, and there is no requirement that N be even. The general case will not be solved separately; rather, it will be transformed into the special case in Section 2.3.

Define $K - k$ new units, $a_1, \ldots, a_{K-k}$, called sinks. They will be matched to $K - k$ of the α's and discarded, so that only k of the α's are matched to β's or γ's. Similarly, define $M - m$ new units, $b_1, \ldots, b_{M-m}$, and $N - n$ new units, $c_1, \ldots, c_{N-n}$. Let $\mathcal{E}$ contain the old units together with the new ones:

$$\mathcal{E} = \{\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_M, \gamma_1, \ldots, \gamma_N, a_1, \ldots, a_{K-k}, b_1, \ldots, b_{M-m}, c_1, \ldots, c_{N-n}\}.$$

Now $\mathcal{E}$ contains $\mathcal{D}$ together with the new units, and $\delta(\cdot)$ is defined and strictly positive for all pairs of units in $\mathcal{D}$. Extend this definition of $\delta(\cdot)$ to $\mathcal{E}$ as follows:

$$\delta\{(w, a_j)\} = \infty \quad \text{for} \quad w \in \mathcal{E} - \mathcal{A},\ \delta\{(\alpha_i, a_j)\} = 0, \quad \text{for} \quad i = 1, \ldots, K,\ j = 1, \ldots, K - k,$$

$$\delta\{(w, b_j)\} = \infty \quad \text{for} \quad w \in \mathcal{E} - \mathcal{B},\ \delta\{(\beta_i, b_j)\} = 0, \quad \text{for} \quad i = 1, \ldots, M,\ j = 1, \ldots, M - m,$$

$$\delta\left\{\left(w, c_j\right)\right\}=\infty \quad \text{for} \quad w \in \mathcal{E}-\mathcal{C}, \delta\left\{\left(\gamma_i, c_j\right)\right\}=0, \quad \text{for} \quad i=1, \ldots, N, j=1, \ldots, N-n.$$

Claim 2. *Let $\mathcal{F}$ be an optimal nonbipartite matching from $\mathcal{E}$, and let $\mathcal{P}$ be the subset of $\mathcal{F}$ that is comprised of pairs that do not contain any of the new units, a_i, b_j, c_u. If $\Delta(\mathcal{F})<\infty$, then $\mathcal{P}$ is an optimal tripartite matching, and if $\Delta(\mathcal{F})=\infty$, then there is no feasible tripartite matching.*

Proof: If $\Delta(\mathcal{F})<\infty$, then each a_j is matched to an α_i, each b_j is matched to a β_i, and each c_j is matched to a γ_i, so that k of the α's are matched to β's or γ's, m of the β's are matched to α's or γ's, and n of the γ's are matched to α's or β's, so $\mathcal{P}$ has the correct (k, m, n). Notice that

$$\left[\begin{array}{lll} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right]\left[\begin{array}{c} P_{\alpha\beta} \\ P_{\alpha\gamma} \\ P_{\beta\gamma} \end{array}\right]=\left[\begin{array}{c} k \\ m \\ n \end{array}\right], \tag{2.1}$$

where the matrix is nonsingular, so having the correct (k, m, n) entails having the correct size $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$. That is, every nonbipartite matching $\mathcal{F}$ from $\mathcal{E}$ with $\Delta(\mathcal{F})<\infty$ yields a $\mathcal{P}$ that is a tripartite matching of the correct size. Conversely, every tripartite matching $\mathcal{P}$ with $\Delta(\mathcal{P})<\infty$ yields a nonbipartite matching $\mathcal{F}$ from $\mathcal{E}$ with $\Delta(\mathcal{F})<\infty$ by pairing α's not matched in $\mathcal{P}$ to a's in $\mathcal{F}$, and so on. Moreover, $\Delta(\mathcal{F})=\Delta(\mathcal{P})$. It follows that $\mathcal{F}$ is an optimal nonbipartite matching of $\mathcal{E}$ with $\Delta(\mathcal{F})<\infty$ if and only if $\mathcal{P}$ is an optimal tripartite matching with $\Delta(\mathcal{P})<\infty$. □

2.5 MATCHING WITH MORE THAN THREE GROUPS

The algorithm as described solves a quite general problem with three groups, and because it is very rare for observational studies to include more than two control groups, the algorithm as described solves the problem as it arises in common practice. Specifically, with three groups, the investigator can specify any pairwise counts $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$—for example, equal counts, $P_{\alpha\beta}=P_{\alpha\gamma}=P_{\beta\gamma}$, for a balanced incomplete block design—and the algorithm will find the optimal match with this choice of $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$. Alternatively, one may use every available unit by setting $k=K$, $m=M$, and $n=N$, use (2.1) to solve uniquely for $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$, and then use the algorithm to obtain the best match that uses all the units. Notice that in this second case, the investigator accepts whatever pairwise counts $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$ happen to be implied by the structure of the data. What is possible with four or more groups?

In the case of three groups, the marginal counts (k, m, n) determine the pairwise counts $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$ by way of (2.1) because the matrix is nonsingular. The algorithm ensures the correct (k, m, n) and hence the correct $\left(P_{\alpha\beta}, P_{\alpha\gamma}, P_{\beta\gamma}\right)$. With four groups, there are four marginal counts and $\binom{4}{2}=6$ pairwise counts, so the analog of (2.1) has four equations in six unknowns, and obtaining the correct marginal counts does not ensure the correct pairwise counts. As a result, the algorithm cannot be used to generate a balanced incomplete block design with four or more groups. With four or more groups, however, the algorithm may still be used to optimally pair match all of the units, by specifying the marginal counts required and applying the algorithm as before, and accepting the pairwise counts that result.

In short, with three groups, the algorithm will (1) find the best balanced incomplete pair matching and (2) find the best pair matching that uses all observations, but for four or more groups, the algorithm will solve (2) but not (1).

3. EXAMPLE: TWO CONTROL GROUPS IN THE MINIMUM WAGE DATA

3.1 FOUR COVARIATES, THREE GROUPS, TWO MATCHED DESIGNS

For illustration, the method will be applied to Card and Krueger's (1994) minimum wage data, discussed in Section 1, creating two simple incomplete block designs, in which three groups are compared in matched pairs. The first design in Section 3.2 is a balanced incomplete block design, closely resembling the corresponding experimental design, whereas the second design in Section 3.3 is unbalanced but uses all restaurants with complete data on the variables we use. We also present in Section 3.4 a brief analysis of the unbalanced design.

Here, the set $\mathcal{A}$ contains the $K = 84$ New Jersey restaurants paying a starting wage of \$4.25 before the wage increase, $\mathcal{B}$ contains the $M = 57$ New Jersey restaurants paying a starting wage $\geq$ \$5.00 before the wage increase, and $\mathcal{C}$ contains the $N = 63$ Pennsylvania restaurants unaffected by New Jersey's wage increase. Recall that economic theory would predict a decline in employment when group $\mathcal{A}$ is compared to groups $\mathcal{B}$ and $\mathcal{C}$, and if New Jersey and Pennsylvania are comparable, employment changes in groups $\mathcal{B}$ and $\mathcal{C}$ should be similar.

We match on four covariates describing the restaurants before the wage increase. One covariate, the chain, had four nominal categories, Burger King, Wendy's, Roy Roger's, and Kentucky Fried Chicken, and it was matched by setting the distance $\delta(\cdot)$ to a large number for pairs of restaurants from different chains. In the balanced incomplete block design, every pair is matched for chain. In the larger unbalanced incomplete block design, 3 of the 102 matched pairs are mismatched for chain because an exact match was not possible. The other three covariates described restaurants before the wage increase: hours open on a weekday, company owned or not, and starting wage before the increase. For restaurants from the same chain, the distance is the Mahalanobis distance using these three covariates (Rubin 1980). By definition, groups $\mathcal{A}$ and $\mathcal{B}$ must differ substantially on one covariate, the starting wage before the increase, since this variable defines the groups; however, this is not a problem.

3.2 DESIGN 1: A BALANCED INCOMPLETE BLOCK DESIGN

The largest possible *balanced* incomplete block design has $k = m = n = 56$ restaurants from each group, and $P_{\alpha\beta} = P_{\alpha\gamma} = P_{\beta\gamma} = 28$ pairs of each type, and we construct the

Table 1. Covariate Balance After Matching: Covariate Means in 3 Groups of 28 Matched Pairs.

Pairing	*Hours open*	*Company owned*	*Starting wage*
NJ \$4.25	15.3	21%	\$4.25
PA	15.2	21%	\$4.39
NJ \$4.25	13.6	39%	\$4.25
NJ $\geq$ \$5.00	14.0	39%	\$5.02
PA	13.9	50%	\$4.93
NJ $\geq$ \$5.00	13.6	46%	\$5.19

best such design. This means that $3 \times 28 \times 2 = 168$ restaurants are included in this design, of 204 available. In Section 3.3 an unbalanced design is constructed that uses all 204 restaurants. Following the procedure in Section 2.4, the set $\mathcal{E}$ contains the $204 = 84 + 57 + 63$ restaurants together with sinks

$$\mathcal{E} = \{\alpha_1, \ldots, \alpha_{84}, \beta_1, \ldots, \beta_{57}, \gamma_1, \ldots, \gamma_{63}, a_1, \ldots, a_{28}, b_1, c_1, \ldots, c_7\}\,.$$

Here, the $K - k = 84 - 56 = 28$ sinks $a_1, \ldots, a_{28}$ are matched to restaurants from $\mathcal{A}$ leaving $k = 56$ other restaurants to be matched to $\mathcal{B}$ and $\mathcal{C}$, and so on. We extend the definition of $\delta(\cdot)$ to $\mathcal{E}$ as described in Section 2.4 and then apply Derigs' algorithm for optimal nonbipartite matching, yielding the best balanced incomplete block design.

The algorithm yields 3 groups of 28 pairs of restaurants, or $3 \times 28 \times 2 = 168$ restaurants in total. The pairs are exactly matched for chain. For the remaining three covariates, Table 1 describes the degree of covariate balance after matching. Each row of the table describes 28 restaurants, and adjacent rows are paired. Overall, the covariate balance achieved by matching is fairly good, except for the deliberate and inevitable difference in starting wages for the two New Jersey groups defined by their starting wages, NJ \$4.25 versus NJ $\geq$ \$5.00. Notice, in particular, that the 28 (NJ \$4.25, PA) pairs describe restaurants paying lower starting wages before the increase, most of which were not company owned, while the 28 (PA, NJ $\geq$ \$5.00) pairs describe restaurants paying higher starting wages before the increase, about half of which were company owned. Although there is comparability within groups of pairs on the four covariates, there is heterogeneity between the three groups of pairs. The matching has been quite effective in producing pairs that appear comparable in terms of the observed covariates.

3.3 DESIGN 2: AN UNBALANCED INCOMPLETE BLOCK DESIGN

The second design matches all 204 restaurants; however, this requires an unbalanced incomplete block design, with more pairs of one type than of another. If all restaurants are to be used, then Equation (2.1) uniquely determines the number of restaurants of each type. Solving Equation (2.1) with $k = 84$, $m = 57$, $n = 63$, yields the number of pairs as $P_{\alpha\beta} = 39$, $P_{\alpha\gamma} = 45$, and $P_{\beta\gamma} = 18$, so there will be 39 pairs in which a New Jersey restaurant paying \$4.25 is matched to a New Jersey restaurant paying $\geq$ \$5, and so on. As

Table 2. Covariate Balance After Matching: Covariate Means in Three Groups of Matched Pairs in an Unbalanced Design.

Pairing	*Hours open*	*Company owned*	*Starting wage*
NJ \$4.25 ($P_{\alpha\gamma} = 45$)	14.8	22%	\$4.25
PA ($P_{\alpha\gamma} = 45$)	14.9	27%	\$4.57
NJ \$4.25 ($P_{\alpha\beta} = 39$)	13.9	36%	\$4.25
NJ $\geq$ \$5.00 ($P_{\alpha\beta} = 39$)	14.1	41%	\$5.08
PA ($P_{\beta\gamma} = 18$)	13.4	67%	\$4.88
NJ $\geq$ \$5.00 ($P_{\beta\gamma} = 18$)	13.4	44%	\$5.18

in Section 2.4, the set $\mathcal{E}$ contains $204 = 84 + 57 + 63$ restaurants without sinks:

$$\mathcal{E} = \{\alpha_1, \ldots, \alpha_{84}, \beta_1, \ldots, \beta_{57}, \gamma_1, \ldots, \gamma_{63}\}.$$

Three of the 102 matched pairs are mismatched for chain, and the remaining 99 pairs are exactly matched for chain. Specifically, three of the $k = 84$ NJ restaurants paying \$4.25 could not be matched for chain; they were all Burger Kings, one being matched to a KFC in Pennsylvania, the other two being matched to Wendy's in New Jersey paying $\geq$ \$5. Table 2 shows the balance within pairs on the other three covariates, where the $P_{\alpha\beta} = 39$ low-wage/high-wage pairs in New Jersey are deliberately mismatched for starting wage. The matching is again quite good, except for the imbalance in company ownership for the third group, (β, γ), of 18 pairs.

3.4 ANALYSIS OF THE UNBALANCED DESIGN

The analysis uses the difference-in-differences method (Angrist and Krueger 1999) within pairs; specifically, the change in full-time equivalent employment (FTE), after-minus-before, is compared in treated and control groups. Table 3 presents an elementary analysis of the unbalanced design in Section 3.3, using the Wilcoxon signed rank statistic and the Hodges-Lehmann point estimate (Lehmann 1998) applied to the treated-minus-control difference in after-minus-before changes in FTE. This analysis would be justified by randomization in a randomized block experiment, but in an observational study such as this, the analysis assumes that the matching has controlled biases due to nonrandom assignment; see Rosenbaum (2002a, sec. 4) where alternative assumptions and analyses are discussed. None of the differences are statistically significant in two-sided 0.05 level tests, and the point estimates have signs opposite those predicted by economic theory. Specifically, the NJ \$4.25 restaurants gained 1.00 more employee than matched PA restaurants and 2.50 more employees than NJ $\geq$ \$5 restaurants; that is, the group predicted to lose workers due to the enforced wage increase had, instead, relative gains rather than losses in employment. The two control groups, PA and NJ $\geq$ \$5, do not differ significantly, so there is little evidence against the comparability of New Jersey and Pennsylvania in terms of changes in FTE at fast food restaurants.

Table 3. Analysis of the Unbalanced Design: Change in FTE Employment.

Pairing	*HL estimate*	*Signed rank p value*
NJ \$4.25 vs PA	1.00	0.66
NJ \$4.25 vs NJ $\geq$ \$5	2.50	0.08
PA vs NJ $\geq$ \$5	-1.25	0.76

The separate analyses in Table 3 may be combined. If restaurants within matched pairs were comparable and the effect of the wage increase was an additive constant for the NJ \$4.25 restaurants, then the 45 NJ \$4.25 versus PA pairs could be combined with the 39 NJ \$4.25 versus NJ $\geq$ \$5 pairs, yielding 84 disjoint pairs, a two-sided significance level of 0.16 from the signed rank statistic, and a Hodges-Lehmman point estimate of 1.75 FTE, which again has a sign opposite that predicted by economic theory.

Other analyses of the design are possible. The effect on employment might be modeled in terms of the magnitude of the effect on wages (Rosenbaum 1999a). The analysis might combine matching with covariance adjustment (Rubin 1979; Rosenbaum 2002b). In observational studies, blocking or matching is intended primarily to control bias (Cochran 1965), whereas in a randomized block experiment, it is simply used to increase efficiency. In the case of an experiment, one could then combine within block and between block information, so that some NJ \$4.25 versus PA comparisons are within the 45 matched pairs and others are between unmatched NJ \$4.25 and PA restaurants; see Wei (1982) for such an analysis.

ACKNOWLEDGMENTS

Supported by NSF Grant SES-0345113.

[Received July 2001. Revised January 2003.]

REFERENCES

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993), *Network Flows: Theory, Algorithms, and Applications*, New York: Prentice Hall.

Angrist, J. D., and Krueger, A. B. (1999), "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics, III*, New York: Elsevier, sec. 23.

Bergstralh, E. J., Kosanke, J. L., and Jacobsen, S. L. (1996), "Software for Optimal Matching in Observational Studies," *Epidemiology*, 7, 331–332; http://www.mayo.edu/hsr/sasmac.html.

Bertsekas, D. P. (1991), *Linear Network Optimization*,Cambridge, MA: MIT Press; http://www.mit.edu:8001//people/dimitrib/home.html.

Campbell, D. (1969), "Artifact and Control," in *Artifact in Behavioral Research*, eds. R. Rosenthal and R. Rosnow, New York: Academic, pp. 351–382.

Card, D., and Krueger, A. (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84, 772–793.

Chang, W. P., Hwang, B., Wang, D., and Wang, J. (1997), "Cytogenetic Effect of Chronic Low-Dose, Low-Dose-Rate Gamma-Radiation in Residents of Irradiated Buildings," *Lancet*, 350, 330–333.

Cochran, W. G. (1965), “The Planning of Observational Studies of Human Populations,” *Journal of the Royal Statistical Society*, Series A, 128, 134–155.

Dehejia, R. H., and Wahba, S. (1999), “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94, 1053–1062.

Derigs, U. (1988), “Solving Non-bipartite Matching Problems via Shortest Path Techniques,” *Annals of Operations Research*, 13, 225–261.

Douglas, D., and Carney, G. (1998), “Exposure to Asphalt or Bitumen Fume and Renal Disease,” *Occupational and Environmental Medicine*, 55, 645–646.

Ghavamian, R., Bergstralh, E. J., Blute, M. L., Slezak, J., and Zincke, H. (1999), “Radical Retropubic Prostatectomy Plus Orchectomy Versus Orchietomy for pTxN+ Prostate Cancer,” *Journal of Urology*, 161, 1223–1227.

Lehmann, E. L. (1998), *Nonparametrics*, New Jersey: Prentice Hall.

Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. R. (2001), “Matching With Doses in an Observational Study of a Media Campaign Against Drug Abuse,” *Journal of the American Statistical Association*, 96, 1245–1253.

Meyer, B. D. (1995), “Natural and Quasi-Experiments in Economics,” *Journal of Business and Economic Statistics*, 13, 151–161.

Nuako, K. W., Ahlquist, D. A., Sandborn, W. J., Mahoney, D. W., Siems, D. M., and Zinsmeister, A. R. (1998), “Primary Sclerosing Cholangitis and Colorectal Carcinoma in Patients With Chronic Ulcerative Colitis,” *Cancer*, 82, 822–826.

Papadimitriou, C., and Steiglitz, K. (1982), *Combinatorial Optimization*, Englewood Cliffs, NJ: Prentice-Hall.

Reinisch, J. M., Sanders, S. A., Mortensen, E. L., and Rubin, D. B. (1995), “In Utero Exposure to Phenobarbital and Intelligence Deficits in Adult Men,” *Journal of the American Medical Association*, 274, 1518–1525.

Rephann, T., and Isserman, A. (1994), “New Highways as Economic Development Tools—An Evaluation Using Quasi-Experimental Matching Methods,” *Regional Science and Urban Economics*, 24, 723–751.

Roghmann, K., and Sodeur, W. (1972), “The Impact of Military Service on Authoritarian Attitudes,” *American Journal of Sociology*, 78, 418–433.

Rosenbaum, P. R. (1987), “The Role of a Second Control Group in an Observational Study” (with discussion), *Statistical Science*, 2, 292–316.

——— (1989), “Optimal Matching in Observational Studies,” *Journal of the American Statistical Association*, 84, 1024–1032.

——— (1999a), “Using Combined Quantile Averages in Matched Observational Studies,” *Applied Statistics*, 48, 63–78.

——— (2002a), *Observational Studies*, New York: Springer.

——— (2002b), “Covariance Adjustment in Randomized Experiments and Observational Studies,” *Statistical Science*, 17, 286-327.

Rosenbaum, P., and Rubin, D. (1985), “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *The American Statistician*, 39, 33–38.

Rosenbaum, P. R., and Silber, J. H. (2001), “Matching and Thick Description in an Observational Study of Mortality After Surgery,” *Biostatistics*, 2, 217–232.

Rubin, D. B. (1973), “The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies,” *Biometrics*, 29, 185–203.

——— (1979), “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies,” *Journal of the American Statistical Association*, 74, 318–328.

——— (1980), “Bias Reduction Using Mahalanobis Metric Matching,” *Biometrics*, 36, 293–298.

Seltser, R., and Sartwell, P. (1965), “The Influence of Occupational Exposure to Radiation on the Mortality of American Radiologists and Other Medical Specialists,” *American Journal of Epidemiology*, 81, 2–22.

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

Silber, J. H., Rosenbaum, P. R., Trudeau, M. E., Even-Shoshan, O., Chen, W., Zhang, X., and Mosher, R. E. (2001), "Multivariate Matching and Bias Reduction in the Surgical Outcomes Study," *Medical Care*, 39, 1048–1064.

Smith, H. (1997), "Matching With Multiple Controls to Estimate Treatment Effects in Observational Studies," *Sociological Methodology*, 27, 325–353.

Smith, A., Kark, J., Cassel, J., and Spears, G. (1977), "Analysis of Prospective Epidemiologic Studies by Minimum Distance Case–Control Matching," *American Journal of Epidemiology*, 105, 567–574.

Snyder, T. L., and Steele, J. M. (1990), "Worst Case Greedy Matching in the Unit d-cube," *Networks*, 20, 779–800.

Warner, D. O., Warner, M. A., Offord, K. P., Schroeder, D. R., Maxson, P., and Scanlon, P. D. (1999), "Airway Obstruction and Perioperative Complications in Smokers Undergoing Abdominal Surgery," *Anesthesiology*, 90, 372–379.

Wei, L. J. (1982), "Interval Estimation of Location Difference with Incomplete Data," *Biometrika*, 69, 249–251.

Wells, K. E., Roberts, C., Daniels, S. M., Hann, D., Clement, V., Reintgen, D., and Cox, C. E. (1997), "Comparison of Psychological Symptoms of Woman Requesting Removal of Breast Implants with Those of Breast Cancer Patients and Healthy Controls," *Plastic and Reconstructive Surgery*, 99, 680–685.

Weston, J. F., and Mansinghka, S. K. (1971), "Tests of the Efficiency Performance of Conglomerate Firms," *Journal of Finance*, 26, 919–936.

Zabin, L. S., Hirsch, M. B., and Emerson, M. R. (1989), "When Urban Adolescents Choose Abortion: Effects on Education, Psychological Status, and Subsequent Pregnancy," *Family Planning Perspectives*, 21, 248–255.