Imposing Minimax and Quantile Constraints on Optimal Matching in Observational Studies

Paul R. Rosenbaum

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania

ABSTRACT

Modern methods construct a matched sample by minimizing the total cost of a flow in a network, finding a pairing of treated and control individuals that minimizes the sum of within-pair covariate distances subject to constraints that ensure distributions of covariates are balanced. In aggregate, these methods work well; however, they can exhibit a lack of interest in a small number of pairs with large covariate distances. Here, a new method is proposed for imposing a minimax constraint on a minimum total distance matching. Such a match minimizes the total within-pair distance subject to various constraints including the constraint that the maximum pair difference is as small as possible. In an example with 1391 matched pairs, this constraint eliminates dozens of pairs with moderately large differences in age, but otherwise exhibits the same excellent covariate balance found without this additional constraint. A minimax constraint eliminates edges in the network, and can improve the worst-case time bound for the performance of the minimum cost flow algorithm, that is, a better match from a practical perspective may take less time to construct. The technique adapts ideas for a different problem, the bottleneck assignment problem, whose sole objective is to minimize the maximum within-pair difference; however, here, that objective becomes a constraint on the minimum cost flow problem. The method generalizes. Rather than constrain the maximum distance, it can constrain an order statistic. Alternatively, the method can minimize the maximum difference in propensity scores, and subject to doing that, minimize the maximum robust Mahalanobis distance. An example from labor economics is used to illustrate. Supplementary materials for this article are available online.

1. Introduction: Avoiding Poor Matches When Optimizing the Typical Match

1.1 Background

In observational studies, matching is used to construct a control group that resembles the treated group in terms of measured covariates. Modern methods solve a discrete optimization problem subject to constraints, minimizing the sum of within-pair covariate distances subject to constraints that ensure covariate balance in marginal distributions; see Section 2 for an example. Minimizing a sum or mean of within-pair differences subject to constraints often produces good matched samples. Nonetheless, a mean over thousands of pairs does little to constrain the worst pairs, so minimizing the mean pair difference is consistent with having dozens of poorly matched pairs. The current article discusses a general algorithm for appending one or more minimax constraints, or constraints on high-order statistics, thereby avoiding poorly matched pairs to the extent that this is feasible given the data available. The new general algorithm in Section 4 permits many useful variations on the theme of constraining the worst to be tolerable before optimizing the mean.

For discussion of optimal matching in observational studies, see Rosenbaum (1989, 2010, Part II), Hansen and Klopfer (2006), Hansen (2007), Stuart (2010), Lu et al. (2011), Yang et al. (2012), Zubizarreta (2012), and Pimentel et al. (2015), and see R packages optmatch, nbpMatching, and rcbalance at CRAN, finebalance archived but working at CRAN, and mipmatch available from Zubizarreta (2012). For applications of optimal matching, see Silber et al. (2013, 2014) and Neuman et al. (2014).

1.2 A Toy Illustration: Three Treated Subjects and Three Controls

Figure 1 is a toy example that will be used to discuss several issues. Figure 1 has three treated subjects, $\tau_1$, $\tau_2$, $\tau_3$, and three controls, $\gamma_1$, $\gamma_2$, $\gamma_3$, and the task is to select one of the $3! = 6$ possible pairings of treated and control subjects. There is a distance between each $\tau_i$ and each $\gamma_j$, reflecting how similar $\tau_i$ and $\gamma_j$ are with respect to measured covariates. A modern distance is the sum of a robust Mahalanobis distance plus a penalty function that prevents poor matches for a propensity score, and this distance is used in the empirical example in Section 2; see Rosenbaum (2010, chap. 8). However, in the toy example in Figure 1, the form of the distance is not a consideration. The dashed lines in Figure 1 are distances of 100, that is, very poor matches. The heavy, solid, horizontal lines are distances of 1. The light solid lines are marked with their distances of 2, 0, and 0.

A greedy or best-first or nearest-available matching algorithm would start with the smallest available distance, perhaps...
Figure 1. Toy example with three treated subjects, $\tau_1$, $\tau_2$, $\tau_3$, and three controls, $\gamma_1$, $\gamma_2$, $\gamma_3$, who may be matched in $3! = 6$ possible ways. The lines or edges connecting treated and control subjects have costs, which measure how similar individuals are in terms of measured covariates. Dashed lines indicate a large distance of 100, heavy solid lines indicate a distance of 1, and light solid lines have costs indicated by their numeric labels of 2, 0, and 0.

Forming the matched pair $(\tau_3, \gamma_2)$ for a cost of 0. That would be a fatal mistake. If a greedy algorithm played chess, it would capture a piece whenever it could, with no thought to what comes next, and it would play terrible chess. In Figure 1, starting with matched pair $(\tau_3, \gamma_2)$ at cost 0, the greedy algorithm would pair $(\tau_1, \gamma_1)$ for a cost of 1, and would be forced to accept $(\tau_2, \gamma_3)$ for a cost of 100, so the match has total distance $0 + 1 + 100 = 101$ and maximum distance 100. If Figure 1 were changed by increasing the cost of the dashed lines, the greedy solution could be made to have both a total and maximum distance that is arbitrarily large, that is, extremely poor, when compared with optimal matching.

In Figure 1, there are two pairings that minimize the total cost over the $3! = 6$ possible pairings. One optimal solution pairs $(\tau_1, \gamma_3)$ for a cost of 2, $(\tau_2, \gamma_2)$ for a cost of 1, and $(\tau_3, \gamma_1)$ for a cost of 0, making a total distance of $2 + 1 + 0 = 3$ and a maximum distance of 2. The other optimal solution pairs $(\tau_1, \gamma_1)$ for a cost of 1, $(\tau_2, \gamma_2)$ for a cost of 1, and $(\tau_3, \gamma_3)$ for a cost of 1, making a total distance of $1 + 1 + 1 = 3$ and a maximum distance of 1. Typical matching algorithms minimize the total distance, perhaps with additional balance constraints, and these typical algorithms would be indifferent between the two optimal solutions just described, because they have the same optimum, that is, the user would simply receive one of the optimal solutions picked arbitrarily. In contrast, the most straightforward application of the algorithm in the current article would have picked the solution $(\tau_1, \gamma_1)$, $(\tau_2, \gamma_2)$, and $(\tau_3, \gamma_3)$, thereby having a maximum distance of 1. Figure 1 is a toy example with three pairs. In problems of realistic size, minimizing the total distance is consistent with having some pairs that have very large distances.

The proposed new algorithm combines elements of two problems: in its simplest form, it minimizes the maximum distance, and subject to the constraint that it does that it minimizes the total distance. In Figure 1, a first step of the algorithm would discover that, among the six possible pairings, the pairing with the smallest maximum distance has a maximum distance of 1. In Figure 1, only one of the six possible pairings has a maximum distance of 1, namely, $(\tau_1, \gamma_1)$, $(\tau_2, \gamma_2)$, and $(\tau_3, \gamma_3)$, that is, there is only one pairing in Figure 1 that minimizes the maximum distance. Typically, in problems of realistic size, there will be a vast number of matchings that minimize the maximum distance, and the algorithm in the current article would pick the best of these, the one that minimizes the total distance subject to the constraint of minimizing the maximum distance.

The algorithm in the current article first discovers that it is possible to match everyone in Figure 1 with a distance of at most 1; then it removes all edges with a distance greater than 1, and minimizes the total distance in the sparser network that remains. In Figure 1, it was immediately obvious that we should avoid the dashed lines if possible, and the new algorithm discovers that it is possible and removes them very early in its activity, also removing the edge with a distance of 2. In a large problem, if one removed all edges with a distance greater than some number $d$, then one might pick the wrong value of $d$, either removing so many edges that no pairing exists, or else tolerating distances that are much larger than needed. The algorithm in
the current article finds the smallest $d$ that permits a match to be constructed. In general, the algorithm in the current article may increase the total distance by constraining the maximum, but the increase is zero in Figure 1 and small in the example in Section 2; moreover, the user may construct both matches and compare them, as in Figure 1 and the example in Section 2.

In Figure 1, the minimax constraint and the total cost refer to the same distance; however, this is not essential. For instance, the minimax constraint might refer to a distance that involves a few key covariates or scores derived from them. Subject to the minimax constraint, one minimizes a total distance that may include many more covariates.

It can happen that even with a minimax constraint, the optimal match must tolerate a few large distances, so the minimax constraint may be insufficiently constraining on its own. In this case, it may be useful to eliminate edges larger than the minimax cut, then additionally impose another constraint on a high quantile, say the 99% quantile; see Section 4.5.

Investigators sometimes match “with replacement,” allowing matched sets to overlap, and Figure 1 illustrates how inefficient that can be. For illustration, suppose that, unknown to the investigator, the matching variables are irrelevant, responses in the treated group are $N(0,1)$ and in the control group are $N(0,1)$, and all responses are independent. Then, for all $6 = 3!$ pair matchings, the treated-minus-control difference in means is Normal with expectation $\gamma$ and variance $1/3 + 1/3 = 2/3$. Matching-with-replacement to all nearest controls in Figure 1 would: (i) form matched sets $(r_1, \gamma_1)$, $(r_2, \gamma_2)$, and $(r_5, \gamma_1, \gamma_2)$, discarding $\gamma_3$, (ii) compute the average over the three sets of the treated-minus-average-control response in each set, (iii) yield an estimate that is Normal with expectation $\gamma$ and variance $1/3 + 1/2 = 5/6$, which is $(5/6)/2/3 = 1.25$ or inflated by 25%. The nine distances in Figure 1 have an average value of $305/9 = 33.89$. The greedy match above has an average within matched-pair distance of $101/3 = 33.67$, or 99.3% of the overall average. The optimal pair match has an average within-pair distance of $3/3 = 1$ or 0.3% of the overall average. Matching-with-replacement has a weighted average within pair distance of $(1 + 1 + 0) = 2/3$ or 0.2% of the average. In other words, matching-with-replacement reduced the within pair distance from 0.3% of the average to 0.2% of the average and increased the variance of the estimator by 25%.

### 1.3 Computational Advantages of Minimax Constraints

An optimal matching can be found using a minimum cost flow algorithm whose worst case time bound involves the number of edges in a network, as reviewed in Section 3.1. The minimax constraint reduces the number of edges, without discarding attractive edges. As shown by Pimentel et al. (2015), use of a sparse network, with fewer edges, makes possible much more effective use of fine balance constraints.

A recent innovation in the algorithmic theory of matching solves the optimal matching problem in a sequence of stages, by peeling away the least attractive edges, thereby simplifying the problem; see Kao et al. (2001). The algorithm in the current article is related in spirit—it too peels away the least attractive edges. However, where Kao et al. solve the original optimal matching problem, the algorithm in the current article changes the objective: it minimizes the total distance subject to the constraint that the maximum distance is small.

### 1.4 Statistical Advantages: Matching and the Investigation of Unmeasured Biases

Matching adjusts for measured covariates, but the central problem in observational studies is potential bias from unmeasured covariates. The homogeneity of matched pairs affects, in three ways, the investigation of potential biases from unmeasured covariates.

Reducing heterogeneity of responses reduces sensitivity to unmeasured biases; see Rosenbaum (2005). For instance, let $Y_i$, be the treated-minus-control difference in outcomes in the $i$th matched pair. If a closer match on key covariates reduced the standard deviation of a Gaussian $Y_i$ by $20\%$ without altering its expectation, the closer match would make the study insensitive to substantially larger unmeasured biases; more precisely, it would increase the design sensitivity. This is illustrated in an observational study in education by Zubizarreta, Paredes, and Rosenbaum (2014), where matching balanced many covariates, but pairs were made homogenous on the variables most predictive of $Y_i$.

Effect modification refers to an interaction between the magnitude of a treatment effect and a measured covariate. To discover effect modification is to discover that an average treatment effect combines a larger and a smaller effect for different subpopulations, and the larger effect is likely to be insensitive to larger unmeasured biases; see Hsu, Small, and Rosenbaum (2013) and Hsu et al. (2015). In looking for effect modifying covariates, it is helpful to have individual pairs that are closely matched for these covariates, so main effects are not confused with interactions.

The most direct way to address unmeasured biases is to measure them. This is sometimes possible by examining a small number of individuals who are closely matched for observed covariates. For instance, electronic medical records might be supplemented by reading the far more detailed paper hospital chart for a dozen pairs, as was done by Rosenbaum and Silber (2001) in a study of mortality after surgery; see also Weller and Barnes (2014, sec. 6). To be useful as a check for unmeasured biases, the small number of examined pairs need to be very closely matched for important observed covariates; see Section 5.1.

### 1.5 Outline: A Practical Illustration; a New Algorithm; Practical Extensions

The article is organized as follows. Section 2 uses an example from labor economics to illustrate the practical issues without reference to network optimization. A conventional match is presented in Section 2.1: it is quite good in aggregate, but contains some poorly matched pairs. This conventional match is improved by adding one minimax constraint in Section 2.2, that is, a mean within-pair distance is minimized among all matched samples that minimize a maximum absolute pair difference. The idea of the algorithm in Section 4 is sketched informally in Section 2.3, explaining its relationship to threshold algorithms.
for the bottleneck assignment problem. Sections 2.4–2.6 discuss three variations on the general theme. For simplicity, the example in Section 2.1 constructs matched pairs, but the discussion in later sections is general, and applies with multiple controls matched to each treated subject, variable controls, the full matching of Rosenbaum (1991) and Hansen and Klopfer (2006), and so on; see Section 3.2 for details. Background material and notation about network optimization and statistical matching are reviewed in Section 3. The new algorithm is discussed in Section 4. The algorithm includes an optional, slightly abstract feature that can be used to form a subset of matched pairs that are matched very closely, and Section 5 discusses practical applications of this feature.

2. Motivating Example from Labor Economics

2.1 A Good Conventional Match With Dozens of Poorly Matched Pairs

In 1989, for older workers with low incomes and long job tenure, Austria substantially lengthened the duration of unemployment benefits and increased the corresponding unemployment payment. In an interesting study, Lalive, Van Ours, and Zweimüller (2006) asked whether these changes in incentives caused workers who lost their jobs to stay out of work longer. In particular, they looked at workers who met the age, income, and tenure criteria for these twice-expanded benefits, comparing workers who lost their jobs before the change in benefits (controls) and workers who lost their job after the change (treated), where only the treated group received the newly increased benefits. Lalive, Van Ours, and Zweimüller (2006) sensibly did a number of related comparisons, but the following discussion will use just one matched comparison as an illustration of certain issues in multivariate matching. In particular, Lalive, Van Ours, and Zweimüller (2006) also looked at other groups of workers for whom the benefits did not change, or changed in more limited ways, and these comparisons would require a few additional matches. The data are available at the web-page for Cahuc, Carrillo, and Zylberberg (2014) and also in the aamatch package that serves as a software supplement to the current article and recreates certain examples.

There were 1391 men in the treated group (i.e., after the benefit change) who were out of work but not temporarily laid off, and each one is matched to one of the 3193 possible control men (i.e., before the benefit change). The matching controlled for age, wage in prior job, an indicator of at least 3 years of work in the past 5 years, whether the job was an apprenticeship, married or not, divorced or not, education in three levels, whether or not the previous job was a blue collar job, a seasonal job (construction or tourism), a manufacturing job. (That is, in terms of the variable names used by Lalive et al., the comparison is of type == "PBD" and RR" and type == "control" for men, sfraue==0, who were not temporarily laid off ein_zus==0. The match controlled for age, nwaage_pj, e3_5, lehre, married, divorced, bc, seasonal, manuf, hiEduc, med_Educ.)

The match, called “match 1,” used several common devices; see Rosenbaum (2010, Part II). The match was the solution to an optimization problem subject to two constraints. First, the match was constrained to be exact for education and seasonal employment, meaning that a seasonal worker with a primary education was always matched to a seasonal worker with a primary education; see match 1 in Table 1. Second, the match was constrained to exhibit near-fine-balance for the 16-level nominal variable in Table 2. A nominal covariate exhibits fine-balance if it has the same marginal distribution in treated and control groups, ignoring who is matched to whom; see Rosenbaum (1989, sec. 3.2, 2010, chap. 10) and Zubizarreta (2012). A variable that is exactly matched, like “seasonal” and “education” in Table 1, is necessarily finely balanced, but a variable can be finely balanced without being exactly matched, because fine balance constrains marginal distributions, not who is paired with whom. Fine balance is not quite achievable with these data: in level 7 of Table 2, there are only 15 controls among all 3193 controls, whereas the treated group has 17 men in level 7. Match 1 exhibits near-fine-balance in the sense that it comes as close as possible to fine balance: it uses all 15 controls from level 7, and has one extra control in each of levels 4 and 11; see Yang et al. (2012) for detailed discussion of matching with near-fine balance. Subject to these two constraints, match 1 minimized the sum of 1391 distances within the 1391 matched pairs. The distances were the sum of two parts: a caliper on an estimated propensity score implemented by a penalty function plus a robust Mahalanobis distance based on all of the covariates in Table 3. For the penalty function and robust Mahalanobis distance, see Rosenbaum (2010, sec. 8).

There are a large number of possible pair matchings, $^{3193}_{1391} \times ^{3193}_{1391}$ to be precise, but match 1 is the best of these in the specific sense of satisfying the two constraints and minimizing the total of 1391 within-pair differences. Match 1 was obtained by finding the minimum cost flow in a certain network; see Section 3. Constraints are distinguished from costs: the constraints are

| Table 1. Frequency counts for the two variables, seasonal employment and education, which were exactly matched. A treated person was always matched to a control with the same value of “seasonal” and “education.” |
|---|---|---|---|
| Covariates | Controls | Match 1 | Match 2 | All |
| | Prior employment | Education | Treated | | | |
| Not seasonal | Primary | 904 | 904 | 904 | 904 | 2024 |
| Seasonal | Primary | 406 | 406 | 406 | 406 | 1063 |
| Not Seasonal | Secondary | 28 | 28 | 28 | 28 | 46 |
| Seasonal | Secondary | 10 | 10 | 10 | 10 | 15 |
| Not Seasonal | Tertiary | 38 | 38 | 38 | 38 | 40 |
| Seasonal | Tertiary | 5 | 5 | 5 | 5 | 5 |
| Total | 1391 | 1391 | 1391 | 1391 | 3193 |
Table 2. Distribution of four binary covariates (BC = prior blue collar job, M = prior manufacturing job, 3/5 = worked ≥ 3 of the last 5 years, A = apprenticeship) in the treated group, two matched control groups, (M-1 or M-2), and all controls. Both match M-1 and M-2 came as close as possible to balancing these distributions— the counts are off by 4 in total for both M-1 and M-2—but unlike Table 1 these matches did not pair people with identical values of these covariates.

| Level | BC | M | 3/5 | A | Treated | M-1 | M-2 | All | Counts | Percents |
|-------|----|---|-----|---|---------|-----|-----|-----|--------|----------|
| 1     | 0  | 0 | 0   | 0 | 10      | 10  | 10  | 23  | 0.7    | 0.7      |
| 2     | 0  | 0 | 1   | 0 | 16      | 6   | 6   | 6   | 0.4    | 0.4      |
| 3     | 0  | 1 | 0   | 1 | 90      | 90  | 90  | 152 | 6.5    | 6.5      |
| 4     | 0  | 1 | 1   | 1 | 71      | 72  | 71  | 117 | 5.1    | 5.1      |
| 5     | 0  | 1 | 0   | 0 | 1      | 1   | 1   | 3   | 0.1    | 0.1      |
| 6     | 0  | 1 | 0   | 1 | 1      | 1   | 1   | 1   | 0.1    | 0.1      |
| 7     | 0  | 1 | 1   | 0 | 17      | 15  | 15  | 15  | 1.2    | 1.1      |
| 8     | 0  | 1 | 1   | 1 | 14      | 14  | 14  | 23  | 1.0    | 1.0      |
| 9     | 1  | 0 | 0   | 0 | 68      | 68  | 68  | 300 | 4.9    | 4.9      |
| 10    | 1  | 0 | 0   | 1 | 27      | 27  | 27  | 77  | 1.9    | 1.9      |
| 11    | 1  | 0 | 1   | 0 | 663     | 664 | 664 | 1557| 47.7   | 47.7     |
| 12    | 1  | 1 | 0   | 1 | 226     | 226 | 227 | 427 | 16.2   | 16.3     |
| 13    | 1  | 1 | 0   | 0 | 16      | 16  | 16  | 43  | 1.2    | 1.2      |
| 14    | 1  | 1 | 0   | 1 | 12      | 12  | 12  | 15  | 0.9    | 0.9      |
| 15    | 1  | 1 | 1   | 0 | 120     | 120 | 120 | 335 | 8.6    | 8.6      |
| 16    | 1  | 1 | 1   | 1 | 49      | 49  | 49  | 99  | 3.5    | 3.5      |
| Total | 1391 | 1391 | 1391 | 3193 | 100.0 | 100.0 | 100.0 | 100.0 |

expressed as part of the node-edge structure of the network, while the costs are the within-pair penalized distances described above.

In many respects and by several conventional standards, match 1 was quite good; see “match 1” in Tables 1–3. In Table 3, prior to matching, all controls had fewer workers with secondary or tertiary education, more blue collar workers, more seasonal workers and other differences, and more controls in level 9 in Table 2, but match 1 has eliminated most of these differences and many others. After matching, the mean treated-minus-control pair difference in age was 0.23 years, and the lower and upper quartiles of the differences in age were −0.39 and 0.63 years.

At the same time, a handful of the 1391 pairs were poorly matched for age; see Figure 2. The group of men in this comparison was defined by eligibility for certain benefits, and in particular ages in this group ranged from 40 to 55. Two of the 1391 pairs had an absolute difference in age of slightly more than 10 years, two others differed by more than 8 years, and so on. This occurs because the match minimized a total distance over 1391 pairs, and that minimum occurred with a few comparatively large differences in age. Could one make the maximum difference in age smaller, perhaps much smaller, without ruining the match? Is so, how much smaller?

2.2 Adding a Minimax Constraint to Prevent Poorly Matched Pairs

Each of the \( \binom{3193}{1391} \) possible matches has a maximum absolute difference in age, so that, in principle, we could compute every match and look at its maximum absolute difference in age. Some of these matches also satisfy the exact and near-fine balance constraints. In principle, if computing time were not an issue, one could: (i) examine each of the \( \binom{3193}{1391} \) possible matches, retaining those that satisfy the constraints, (ii) among the retained matches, determine the smallest possible maximum absolute difference in age, (iii) from the subset of matches that both satisfy the constraints and exhibit the smallest possible maximum difference in age, pick the best one, that is a match that minimizes the total of all of the within-pair penalized robust Mahalanobis distances. Of course, steps (i)–(iii) are not computationally feasible if implemented by the exhaustive search just described; however, the algorithm in Section 4 finds in polynomial time the same matched sample. Indeed, if implemented carefully, the algorithm in Section 4 can be faster than the minimum cost flow algorithm without a minimax constraint, because the constraint eliminates many edges in the network, a key aspect of the worst case time bound for many algorithms.

Table 3. Covariate means or percents in the treated group, two matched control groups, and the control group prior to matching. In terms of means, the two matched control groups look similar.

|                  | Treated | Match 1 | Match 2 | All controls |
|------------------|---------|---------|---------|--------------|
| Sample size      | 1391    | 1391    | 1391    | 3193         |
| Secondary edu %  | 2.7     | 2.7     | 2.7     | 2.7          |
| Tertiary edu %   | 3.1     | 3.1     | 3.1     | 3.1          |
| Seasonal job %   | 30.3    | 30.3    | 30.3    | 33.9         |
| Blue collar %    | 84.9    | 85.0    | 85.0    | 89.4         |
| Worked 3 of last 5 years % | 89.9 | 89.9 | 89.9 | 85.3 |
| Manufacturing job % | 16.5 | 16.4 | 16.4 | 16.7 |
| Apprenticeship  | 29.2    | 29.3    | 29.3    | 24.0         |
| Married %        | 62.0    | 62.6    | 62.9    | 64.6         |
| Divorced %       | 15.7    | 15.5    | 15.5    | 13.0         |
| Age (mean)       | 46.4    | 46.2    | 46.3    | 46.6         |
| Prior wage (mean)| 10,114  | 10,137  | 10,124  | 10,374       |
Figure 2. Treated-minus-control pair differences in age in matched samples M-1 and M-2, each containing 1391 matched pairs. The dashed lines are at $\pm 4.261$ years: every difference in match M-2 is inside the dashed lines, and there exists no feasible match that makes all of the differences fit within a narrower band centered at 0. That is, match M-2 satisfies the minimax constraint: it minimizes the maximum absolute difference in age achievable by any that exactly matches “education” and “seasonal” in Table 1 and exhibits near-fine balance for the 16 categories in Table 2.

It turns out that there exists a match that respects the constraints in Tables 1 and 2 with a maximum absolute difference in age within pairs that is less than 4.261 years, but there is no match of the $\binom{3193}{1391} \times 1391!$ possible matches that satisfies the constraints and has a maximum absolute difference in age strictly less than 4.26 years. Showing this takes a tolerable amount of computation. Once one knows this, an additional constraint is imposed on the match, demanding that the maximum absolute difference in age be less than 4.261 years, and this produces match 2 in Tables 1–3. Using the same definition of distance as match 1, match 2 minimizes the total of 1391 within-pair distances subject to the two constraints required of match 1, namely, the constraints in Tables 1 and 2, plus one additional constraint, namely, that it minimizes the maximum absolute difference in age. Match 2 looks similar to match 1, exhibiting an exact match for “education” and “seasonal” in Table 1, a minimal deviation from fine balance of the same size in Table 2, albeit involving levels 11 and 12 rather than levels 4 and 11, and similar covariate balance in Table 3. However, as seen in Figure 2, the maximum absolute difference in age is 4.261 years in match 2, the smallest possible maximum absolute difference. Match 1 and match 2 each include 1391 controls, but they differ in 43 controls and they pair individuals differently. However, match 1 has 62 pairs with an absolute difference in age of more than 4.261 years, but match 2 has no such pairs.

2.3 Relationship Between Minimax Constraints and the Bottleneck Assignment Problem

Match 2 is the simplest example of imposing a minimax constraint on a matching algorithm that minimizes a total within-pair distance subject to additional constraints. The new algorithm in Section 4 is built from existing tools used to solve the bottleneck assignment problem, a matching problem with a single objective, namely, finding a match to minimize the maximum cost or distance; see, for instance, Garfinkel (1971) and Burkard, Dell’Amico, and Martello (2009). The bottleneck assignment problem is relevant to some problems in operations research where the cost of a subtask is the time to complete that subtask, and the whole task is complete when all subtasks are complete; then minimizing the maximum time to complete the subtasks also minimizes the time to complete the whole task. In statistical matching problems, minimizing the maximum distance is not a reasonable objective—we want the typical distance to be very small—but minimizing the maximum distance is a reasonable additional constraint, for example, in avoiding the pointlessly large differences in age in a few dozen pairs in match 1.

Garfinkel (1971) solved the unconstrained bottleneck assignment problem with a so-called threshold algorithm: (a) one picks a threshold $\vartheta$ and tries to match so all absolute age differences are at most $\vartheta$; (b) if (a) is possible, one decreases $\vartheta$ and
repeats (a); if (a) is impossible, one increases \( \theta \) and repeats (a). A binary search quickly narrows the possible values of \( \theta \). Subproblem (a) is a simpler matching problem with a better computational complexity than the general minimum total cost problem, so several iterations of (a) are often not very costly. When used to set a constraint in the minimum total cost algorithm in Section 4, knowledge of Garfinkel’s threshold \( \theta \) permits many edges to be removed from the network—those with age differences larger than \( \theta \)—and this improves the worse case time bound for the revised minimum cost flow problem. In other words, if we only consider candidate pairs that have an absolute age difference less than 4.261 years, then we drastically reduce the number of candidate pairings that need to be considered, and a threshold algorithm has found 4.261 years to the minimum possible value.

### 2.4 Constraining Several Order Statistics in Sequence, Not Just the Maximum

Match 3, not shown in the tables, also satisfied the constraints in Tables 1 and 2, but it imposed two constraints on the absolute difference in age: (i) it forced the largest absolute difference in age to be less than 4.261 years, and (ii) it forced the second largest absolute difference in age to be less than 3.072 years, the smallest possible value. That is, in match 3, the constraints in Tables 1 and 2 are satisfied, all of the absolute differences in age are less than 4.261 years, and only one difference is larger than 3.072 years; moreover, no match that satisfies the constraints yields smaller values for these two largest order statistics. These constraints are understood sequentially: 4.261 is the smallest maximum difference among matches that satisfy the constraints in Tables 1 and 2, whereas 3.072 is the smallest maximum second-order statistic among matches that both satisfy the constraints in Tables 1 and 2 and minimize the maximum absolute difference. Arguably, match 3 is not much better than match 2, because 3.072 years is not a big improvement over 4.261 years. Match 3 is mentioned solely to illustrate a feature of the algorithm in Section 4. One could minimize the maximum of another order statistic, say the 95% quantile, in a similar manner.

### 2.5 Minimizing the Maximum Mahalanobis Distance or Propensity Score Difference

Although match 2 minimized the maximum absolute within-pair difference in a single covariate, here age, the minimax constraint need not refer to a single covariate. Match 4, described in Table 4, resembles Matches 1–3 in being exact for “education” and “seasonal,” and in exhibiting near-fine balance for four variables in Table 2; however, match 4 imposes a constraint of minimizing the maximum of the robust Mahalanobis distances based on 10 covariates without penalties, before going on to minimize the sum of penalized distances. The frequencies for Match 4 corresponding to Table 2 are identical to the frequencies for Match 1, and the same is true for Table 1. The maximum robust Mahalanobis distance of 14.49 in Table 4 is the smallest possible maximum. Additionally, the distribution of matched distances in Table 4 is much smaller than the distribution before matching.

In parallel with Section 2.4, one can impose a sequence of minimax constraints with different objectives. For instance, replacing age by the propensity score, one could constrain the match to minimize the maximum absolute difference in propensity scores, then add the further constraint that among matches that minimize the maximum absolute difference in propensity scores, the match must minimize the maximum robust Mahalanobis distance.

### 2.6 Minimix Constraints Within Exact Match Strata

Some of the restrictions in Table 1 are very restrictive. Consider the first row, workers not in seasonal employment with a primary education. Ignoring the fine balance constraints in Table 2, the workers in the first row can be made into a matched sample in \( \binom{504}{2024} \times 904! \) ways, so there are many choices of who and how to match, thereby reducing within-pair robust Mahalanobis distances. In contrast, in the last row of Table 1—seasonal workers with tertiary education—there are only \( \binom{\frac{504}{2}}{2} \times 5! = 120 \) ways to pair the workers. Counting the number of matched samples that also respect the fine balance constraints in Table 2 is much more difficult because these constraints aggregate the exact match categories in Table 1. Not surprisingly, the largest robust Mahalanobis distance in match 4 in Table 4, namely 14.49, is between two seasonal workers with tertiary education, a small stratum with few choices. Perhaps we should impose a tighter constraint on the robust Mahalanobis distance in the first row of Table 1 than the constraint of 14.49 required for the last row of Table 1. We cannot do better than 14.49 overall, because we cannot do better than 14.49 in the last row of Table 1, but we could have different requirements for different strata.

Specifically, the method in Section 4 may be used to minimize the maximum robust Mahalanobis distance yielding, as in Section 2.5, 14.49 in the sixth stratum in Table 1. With that constraint in place, it can minimize the maximum robust Mahalanobis distance over the union of the five remaining strata. Optionally, this could continue with 4 strata, then 3, and so on. As in the case of the second largest order statistic in Section 2.4, once the maximum is tolerably small, the later steps in this process may not be worth pursuing.

### 3. Notation and Review

#### 3.1 Review: Minimum Cost Flow in a Network

A directed graph is a finite set of nodes, \( \mathcal{N} = \{v_1, \ldots, v_N\} \), together with a finite set \( E \) of directed edges consisting of

|                | Min | 25% | 50% | 75% | 90% | 95%  | 99%  | Max |
|----------------|-----|-----|-----|-----|-----|------|------|-----|
| Before matching| 0.00| 7.70| 10.85| 14.27|17.66|19.80 |23.94|38.63|
| Matched        | 0.00| 0.01| 0.03| 0.16|2.68 |5.02  |8.88 |14.49|
ordered pairs of nodes, so \( E \subseteq \mathcal{N} \times \mathcal{N} \). For a finite set \( \mathcal{S} \) write \(|S|\) for the number of elements of \( \mathcal{S} \), so \(|\mathcal{N}| = N \) and \(|E| \leq N^2\).

A directed graph is drawn on paper by representing each node \( v \in \mathcal{N} \) by a point and each edge \((v, v') \in E\) by an arrow with tail at \( v \) and head at \( v' \). A directed graph is acyclic if there are no cycles, that is, there are no sequences \((v, v'), (v', v''), \ldots, (v^*, v^{**})\) of edges in \( E \) such that \( v = v^{**} \).

A network is an acyclic directed graph \((\mathcal{N}, E)\) such that each node \( v \in \mathcal{N} \) has an integer demand, \( d(v) \in \{\ldots, -2, -1, 0, 1, 2, \ldots\\} \), and each edge \( e \in E \) has a nonnegative real cost, \( p(e) \geq 0 \), and a nonnegative integer capacity \( q(e) \in \{0, 1, 2, \ldots\} \). A flow \( f(\cdot) \) is a function assigning a nonnegative integer \( f(e) \) to each edge \( e \in E \), that is, \( f : E \to \{0, 1, 2, \ldots\} \). A flow respects the capacity constraints if

\[
0 \leq f(e) \leq q(e) \quad \text{for all } e \in E. \tag{1}
\]

A flow meets the demand \( d(v) \) if the total flow into node \( v \) from edges \((v', v) \in E\) minus the total flow out from \( v \) on edges \((v, v') \in E\) equals \( d(v) \), that is, if

\[
d(v) = \sum_{(v', v) \in E} f((v', v)) - \sum_{(v, v') \in E} f((v, v')) \quad \text{for each } v \in \mathcal{N}. \tag{2}
\]

A flow that satisfies both (1) and (2) is feasible. Let \( \mathcal{F} \) be the set containing all the feasible flows \( f \), noting that \( \mathcal{F} \) is empty if no feasible flow exists. The cost \( p_f \) of a feasible flow is \( p_f = \sum_{e \in E} f(e) \cdot p(e) \). If \( \mathcal{F} \neq \emptyset \), then a feasible flow \( f \in \mathcal{F} \) has minimum cost if \( p_f \leq p_f' \) for all feasible flows \( f' \in \mathcal{F} \). As discussed in Section 3.2, many statistical matching problems are solved as minimum cost flow problems in a network; however, other approaches are possible, as discussed by Zubizarreta (2012).

The minimum cost flow problem has several algorithmic solutions with attractive worst-case polynomial time bounds. In practice, this means that large problems can be solved reasonably quickly. Some algorithms have worst-case time bounds that involve only the size of network, \(|\mathcal{N}| \) and \(|E|\), whereas others additionally involve the costs, \( p(\cdot) \), capacities \( q(\cdot) \), or demands \( d(\cdot) \); see Bang-Jensen and Gutin (2009, Section 4.10) for discussion of several algorithms and their worst-case bounds. One good solution is made conveniently available inside R by loading the \texttt{optmatch} package: it is the Fortran code for Relax IV of Bertsekas and Tseng (1988). The calculations in the current article used Relax IV.

A flow vector, \( f(e), e \in E \), is a vector of dimension \(|E|\), and the minimum cost flow problem is a search for an optimal flow vector. The worst case time bounds for good algorithms for the minimum cost flow problem invariably involve \(|E|\), with a much better, much lower time guarantee if \(|E|\) is smaller. A single min-max constraint of the type in Section 2.2 can be implemented by removing high-cost edges (or equivalently large distance pairings) from \( E \), often making \(|E|\) much smaller. In other words, in Section 2.2, changing the problem to disallow large pair differences in age produces a new problem that can be solved more quickly. A sequence of minmax constraints of the type in Section 2.6 can be implemented as a sequence of removals of edges from \( E \), so each step in this sequence has a lower \(|E|\) than the previous step. In practical terms, if we do not adhere to the single-minded pursuit of the smallest total within-pair distance, \( p_f = \sum_{e \in E} f(e) \cdot p(e) \), but rather wish to pursue that objective with the added constraint that no single matched pair has an exceptionally large distance, then we are likely to receive the additional benefit of an improved worst-case time bound by virtue of making \(|E|\) smaller.

### 3.2 Networks Used in Multivariate Matching

The networks used for matching in observational studies have a particular structure. There is a set \( \mathcal{T} = \{t_1, \ldots, t_T\} \) of treated nodes, a set \( \mathcal{C} = \{\gamma_1, \ldots, \gamma_C\} \) of potential control nodes with \( C \geq T \), and a set \( \mathcal{A} = \{a_1, \ldots, a_A\} \) of additional nodes, where \( \mathcal{T}, \mathcal{C}, \mathcal{A} \) are disjoint, and \( \mathcal{N} = \mathcal{T} \cup \mathcal{C} \cup \mathcal{A} \). In Section 2, \( \mathcal{T} \) contains the \( T = 1391 \) treated individuals, \( \mathcal{C} \) contains the \( C = 3193 \) potential controls, and \( \mathcal{A} = \{a_1, \ldots, a_{16}, a_{17}\} \) contains 16 additional nodes, \( a_1, \ldots, a_{16}, \) for the 16 fine balance strata in Table 2 plus one more node \( a_{17} \) that controls the deviation from fine balance. If fine balance were achievable in this example, then node \( a_{17} \) would not be needed. In Section 2, the additional nodes force the near fine balance in Table 2 without pairing people with identical strata. By varying the structure of the additional nodes \( \mathcal{A} \), a variety of balancing or matching properties may be produced. See figures 1 and 2 in Rosenbaum (1989), figure 2 in Rosenbaum (1991), and figure 1 in Pimentel et al. (2015) for depictions of some of the many possible matching networks with additional nodes \( \mathcal{A} \). The new techniques described in the current article will work with any structure for \( \mathcal{A} \), so various structures are mentioned but no one structure is assumed.

There is an edge \( e = (t_i, \gamma_j) \in E \) if \( t_i \) can be matched to \( \gamma_j \). For example, in Table 1, the 5/1391 treated men with prior seasonal employment and tertiary education can only be matched to the 5/3193 control men with prior seasonal employment and tertiary education, so each such treated man \( t_i \) has only five edges in \( E \). In contrast, in Table 1, the 406 treated men with seasonal employment and primary education each have 1063 edges in \( E \), making \( 406 \times 1063 = 431,578 \) in total for seasonal workers with primary education. The cost \( p(e) \) for \( e = (t_i, \gamma_j) \in E \) is the covariate distance between \( t_i \) and \( \gamma_j \). In match 1 in Section 2, the covariate distance between \( t_i \) and \( \gamma_j \) was the sum of the robust Mahalanobis distance between the covariates for \( t_i \) and \( \gamma_j \) plus a penalty if \( t_i \) and \( \gamma_j \) differed on the propensity score by more than the caliper allowed. For each treated-control edge \( e = (t_i, \gamma_j) \in E \), the capacity is \( q(e) = 1 \). For a feasible flow \( f \), if \( e = (t_i, \gamma_j) \in E \), then treated unit \( t_i \) is matched to control \( \gamma_j \) if \( f(e) = 1 \); otherwise, if \( f(e) = 0 \), then \( t_i \) is not matched to \( \gamma_j \).

In the example in Section 2, there is an edge \( e = (\gamma_j, a_{16}) \in E \), connecting each control \( \gamma_j \) to the one fine balance level, \( a_1, \ldots, a_{16} \), that contains \( \gamma_j \); see Table 2. In the example in Section 2, but not in all applications, the cost \( p(e) \) for \( e = (\gamma_j, a_{16}) \in E \) is \( p(e) = 0 \) and the capacity is \( q(e) = 1 \). Also, in this example, each of these 16 fine balance levels, \( a_1, \ldots, a_{16}, \) is connected by an edge to \( a_{17} \) with cost \( p(a_{16}, a_{17}) = 0 \) and capacity \( q(a_{16}, a_{17}) = 2 \). The node \( a_{17} \) will allow at most a treated-minus-control difference in frequencies of 2 in any of the 16 levels in Table 2.

In the example in Section 2, but not in all applications, the demands have the following structure. Every treated unit \( t_i \in \mathcal{T} \) has demand \( d(t_i) = -1 \), meaning that it issues one unit of flow,
and hence $\tau$ must be matched to some control $\gamma$. The demand for every control $\gamma \in C$ is $d(\gamma) = 0$ meaning that $\gamma$ passes on all the flow it receives from treated units $\tau_i$. Because each control $\gamma_i$ is the tail of just one edge $e = (\gamma_i, \alpha_a) \in E$ with capacity $q(e) = 1$, it follows that $\gamma_i$ can receive only one unit of flow from treated units $\tau_i$, so $\gamma_i$ can be matched to only one treated unit. The demands $d(\alpha_a)$ for the 17 additional nodes are determined by examining two columns in Table 2, specifically the counts for treated subjects and all controls. As noted in Section 2, the number of controls in inadequate in level 7 in Table 2 by a count of 7 and the deviation from fine balance would have to be larger than $d(\alpha) = 2$ in Table 2. Minor adjustments permit a match with variable numbers of controls. Keele, Pimentel, and Yoon (2015) set the required number $-d(\tau_i)$ of controls for $\tau_i$ based on their "entire number," a quantity calculated from the propensity score value for individual $\tau_i$. In the hierarchical structure used by Pimentel et al. (2015), some $\alpha_a \in A$ are connected to other $\alpha_d \in A$, where $(\alpha_a, \alpha_d) \in E$ means that the stratum represented by $\alpha_a$ subdivides the coarser stratum represented by $\alpha_d$.

4. Minimizing Total Distance While Constraining Maxima or Quantiles

4.1 Each Possible Match Yields Different Order Statistics

Let $I \subseteq T$ be an interesting subset of the treated subjects, typically all of them, $I \subseteq T$. See Section 5 for discussion of $I \neq T$. For each edge $e = (\tau, \gamma) \in E$ with $\tau \in I$ and $\gamma \in C$ there is a nonnegative number $h_{\tau, \gamma} \geq 0$. In Section 2.2, $h_{\tau, \gamma}$ was the absolute difference in age between $\tau$ and $\gamma$, whereas in Section 2.5 it was a robust Mahalanobis distance. For any feasible flow $f$, there are $L = -\sum_{\tau \in I} d(\tau)$ edges $e = (\tau, \gamma) \in E$ with $\tau \in I$ and $f(\tau, \gamma) = 1$ for $\tau \in I$, that is, there are $L$ controls $\gamma$ matched to the treated subjects $\tau \in I$. For pair matching with $I = T$, there are $L = T$ such edges, one for each matched pair; whereas, with $I = T$ and matching with two controls there are $L = 2T$ such edges. For a given feasible flow, sort the $L$ quantities $h_{(\tau, \gamma)}$ with $\tau \in I$ and $f(\tau, \gamma) = 1$ into nondecreasing order, as $h_{f,1} \leq h_{f,2} \leq \cdots \leq h_{f,L}$, thereby yielding the order statistics for $h_{\tau, \gamma}$ in the match $f$. To emphasize, $h_{f, \tau}$ is different for each match $f$.

First, consider the maximum, $h_{f, L}$. In Section 2.2, $h_{f, L}$ is the maximum absolute difference in age among the $T = L = 1391$ matched pairs. In match 1, $h_{f, L}$ was more than 10 years, but $h_{f, L}$ achieved its minimum possible value of 4.261 for match 2.

In general, if $F \neq \emptyset$, define $h_{\min, f} = \min_{f \in F} h_{f, L}$, so $h_{\min, f}$ is the smallest achievable value of the $\ell$th-order statistic; otherwise, if $F = \emptyset$, then define $h_{\min, f} = \infty$. In Section 2.2, $h_{\min, f} = 4.261$ years. Define $\mathcal{F}_{\min, f} \subseteq F$ to be the subset of feasible flows that achieve the minimum value of $h_{f, L}$, that is, $\mathcal{F}_{\min, f} = \{f \in F : h_{f, L} \leq h_{\min, f}\}$. In Section 2.2, many feasible flows $f$ minimize the maximum absolute difference in age, having $h_{f, L} = h_{\min, f} = 4.261$ years, and match 2 is the best of these in the sense of minimizing the total $p_f$ of the $T = 1391$ robust Mahalanobis covariate distances within matched pairs.

It is important to notice that if $F \neq \emptyset$, then $\mathcal{F}_{\min, f} \neq \emptyset$. In words, whenever there is a feasible match without constraining the order statistic, there is also a feasible match that satisfies the constraint $h_{f, \tau} \leq h_{\min, f}$. The reason for this is that the constraining value, $h_{\min, f}$, was defined in such a way that it was achieved by at least one flow $f \in F$. An attraction of constraints on order statistics is that they never create infeasibility. In contrast, in Section 2.2, one had added the constraint that every pair differ in age by at most $k$ years, then the problem would be feasible for some $k$, infeasible for other $k$, and it would not be obvious by inspecting the distribution of age values of $k$ yield feasibility given the other constraints in Tables 1 and 2. By solving the minimax problem, we know that the match is feasible for $k \geq 4.261$ years and infeasible for $k < 4.262$ years.

4.2 Imposing a Minimax Constraint on a Minimum Total Distance Match

Problem A imposes a single minimax constraint on the minimum cost flow problem. Problem A seeks a feasible flow $f$ such that the total cost $p_f$ of this flow is minimal among all feasible flows $f^*$ with $h_{f^*, L} = h_{\min, f^*}$, that is, among all feasible $f^*$ that minimize the maximum $h_{(\tau, \gamma)}$ for matched pairs with $\tau \in I$.

Problem A (Adding a minimax constraint): Let $F$ be the set of feasible flows for some matching network with the structure $(T \cup C \cup A, E)$ in Section 3.2, let $h_{\min, f} = \min_{f \in F} h_{f, L}$, and let $\mathcal{F}_{\min, f} = \{f \in F : h_{f, L} \leq h_{\min, f}\}$. Find an $f \in \mathcal{F}_{\min, f}$ such that $p_f \leq p_{f^*}$ for all $f^* \in \mathcal{F}_{\min, f}$ or determine that the problem is infeasible in the sense that $F = \mathcal{F}_{\min, f} = \emptyset$.

Problem A was solved in Section 2.2 with $h_{(\tau, \gamma)}$ equal to the absolute difference in age.
4.3 Iterating Minimax Constraints

Problem A can be iterated to impose a sequence of minimax constraints, each new constraint further constraining the previous problem. This is done as follows. Suppose we have solved a problem with the form of Problem A. If this problem is infeasible, if \( F = \emptyset \), then we have determined the number \( h_{\text{min}, L} \). Define a new network \(( T \cup C \cup A, E^* )\) where \( E^* \) consists of all edges of \( E \) except those \(( \tau, \gamma ) \in E \) with \( \tau \in I, \gamma \in C, \) and \( h_{\tau, \gamma} > h_{\text{min}, L} \). By the definition of \( h_{\text{min}, L} \), the set, say \( F^* \), of feasible flows for this new network is not empty because \( F \neq \emptyset \) and a flow \( f \in F \) achieved \( h_{f, L} = h_{\text{min}, L} \). Continuing, if the original problem is feasible with \( F \neq \emptyset \), and if we have another measure \( h^* \), then we may solve a new problem with the form of Problem A but with network \(( T \cup C \cup A, E^* )\) and constraining variable \( h^* \), and this new problem is also feasible, and its solution \( f \) is constrained to minimize \( h_{f, L} \), and constrained to minimize \( h^* \) among flows that minimize \( h_{f, L} \). Because \( |E^*| \leq |E| \), iterates of Problem A have improved worst-case time bounds; see Section 3.1.

For example, one can minimize the maximum absolute difference \( h_\epsilon \) in propensity scores, and subject to that constraint, minimize the maximum robust Mahalanobis distance \( h^* \), as mentioned in Section 2.5. Alternatively, with exact match stratia in Section 2.6, one can minimize the maximum robust Mahalanobis distance \( h_\epsilon \), discover that the maximum was achieved in, say, stratum 6, then set \( h^*_6 = 0 \) for treated \( \tau \) and control \( \gamma \) subjects in stratum 6 and set \( h^*_\tau = h_\epsilon \) otherwise. The next iterate will minimize the maximum robust Mahalanobis distance in all strata other than stratum 6 while forcing stratum 6 to continue to minimize its maximum.

4.4 Threshold Approach to Imposing a Minimax Constraint

Corollary 1 in Section 4.5 shows that Algorithm A below solves Problem A. Step 3a resembles Garfinkel’s (1971) threshold algorithm for the bottleneck assignment problem: it is a binary search to determine \( h_{\text{min}, L} \), whereupon a minimum cost flow problem is solved in Step 4a with a reduced set of edges \( E \).

In Step 3a, the edge set \( E \) gradually shrinks to include fewer and fewer edges. In Step 4a, the edge set \( E \) has been reduced as much as possible. For example, in Section 2.2, had one started with a guess of \( H_{\text{mid}} = 5 \) years, then all edges with an absolute difference in age of more than 5 years would have been removed at the end of the first iteration of Step 3a.

The algorithm maintains the inequality \( H_{\text{low}} \leq h_{\text{min}, L} \leq H_{\text{high}} \), and terminates when \( H_{\text{low}} \leq h_{\text{min}, L} \leq H_{\text{high}} < H_{\text{low}} + \delta \). With the value of \( \delta \) from Step 1a, there is only one value \( h_\epsilon \) in the interval \([ H_{\text{low}}, H_{\text{low}} + \delta ]\), so that value is \( h_{\text{min}, L} \). If one uses a larger value of \( \delta \) in Step 1a, then one is not strictly minimizing the maximum \( h_{f, L} \) in Step 4a, but no pair \( (\tau, \gamma) \) will be produced by step 4a in which \( h_\epsilon > h_{\text{min}, L} + \delta \). A larger value of \( \delta \) means fewer iterates of Step 3a, and perhaps extreme precision in this step is of little practical importance. For example, in Section 2.2, taking \( \delta = 1/2 \) would mean that the maximum absolute difference in age would exceed the minimum possible value, namely, 4.26 years, by at most half a year.

4.5 Threshold Approach to Imposing a Quantile Constraint

Minimizing a quantile other than the maximum can be approached in a similar way, but does not permit further elimination of edges from \( E \). In practice, one might first eliminate all edges with costs greater than the minimax constraint, thereby obtaining a sparse network and minimizing the maximum distance, then impose an additional quantile constraint to further reduce the number of large distances.

Problem B (Adding a quantile constraint): For a fixed \( \ell \in \{1, \ldots, L\} \), let \( F \) be the set of feasible flows for some matching network with structure in Section 3.2, let \( h_{\text{min}, \epsilon} = \min_{f \in F} h_{f, \epsilon} \), and let \( F_{\text{min}, \epsilon} = \{ f \in F : h_{f, \epsilon} \geq h_{\text{min}, \epsilon} \} \). Find an \( f \in F_{\text{min}, \epsilon} \) such that \( p_f \leq p_{f^*} \) for all \( f^* \in F_{\text{min}, \epsilon} \) or determine that the problem is infeasible in the sense that \( F_{\text{min}, \epsilon} = \emptyset \).
Proposition 1. For sufficiently large $\beta$, Algorithm B solves Problem B.

Proof. If the problem is infeasible, this is determined in the first iteration of Step 3, and as noted in Section 4.1, this means the problem was infeasible before adding the constraint on the $\ell$-order statistic. So suppose Problem A is feasible, $\mathcal{F}_{\min,\ell} \neq \emptyset$. In Step 3, if $\sum_{e \in E} \tilde{p}(e) f(e) \leq L - \ell$ then $f$ has fewer than $L - \ell$ edges $e = (\tau, \gamma) \in \mathcal{E}$ with $\tau \in \mathcal{I}$ such that $h_e > H_{\text{mid}}$, so $h_{\min,\ell} \leq h_{\text{mid}}$. In Step 3, for the minimum cost flow $f$, the total cost $\sum_{e \in E} \tilde{p}(e) f(e)$ is the minimum number of edges $e = (\tau, \gamma) \in \mathcal{E}$ with $\tau \in \mathcal{I}$ such that $h_e > H_{\text{mid}}$ and if that minimum is greater than $L - \ell$, then $h_{\min,\ell} > H_{\text{mid}}$. So the algorithm maintains $H_{\text{low}} \leq h_{\min,\ell} \leq H_{\text{high}}$ as $H_{\text{high}} - H_{\text{low}}$ diminishes toward 0 and eventually falls below $\delta$. Once $H_{\text{high}} - H_{\text{low}} < \delta$, the real interval $[H_{\text{low}}, H_{\text{high}}]$ intersects the values of $h_e$ in a single number and that number is $h_{\min,\ell}$; furthermore, $h_e > H_{\text{high}}$ if and only if $h_e > h_{\min,\ell}$. In Step 4, the minimum cost flow $\tilde{p}_f$ is $\sum_{e \in E} p(e) f(e) + B \beta$ where $B$ is the number of edges $e = (\tau, \gamma) \in \mathcal{E}$ with $\tau \in \mathcal{I}$ such that $h_e > H_{\text{high}}$ and $f(e) = 1$. For sufficiently large $\beta$: (i) the minimum cost flow $\tilde{p}_f$ must have the minimum possible value of $B$, (ii) the minimum value of $B$ is attained only by the flows in $\mathcal{F}_{\min,\ell}$, and (iii) among flows in $\mathcal{F}_{\min,\ell}$, the minimum cost flow minimizes $\sum_{e \in E} p(e) f(e)$, as required to solve Problem B.

Corollary 1. Algorithm A solves Problem A.

Proof. In the proof of Proposition 1, if $\ell = L$, then the penalty $\beta$ forces no flow to pass along edges $e = (\tau, \gamma) \in \mathcal{E}$ with $\tau \in \mathcal{I}$ such that $h_e > H_{\text{min,\ell}}$, so these edges could be removed from $\mathcal{E}$ without changing the optimal flow $f$ produced by Algorithm B. However, if these edges are removed, Algorithm B becomes Algorithm A.

5. Using $\mathcal{I}$ to Obtain a Subset of Controls Who Are Very Closely Matched

The methods in Section 4 constrained a maximum or an order statistic of $h_e$ for pairs $(\tau, \gamma) \in \mathcal{E}$ with $\tau \in \mathcal{I}$, but all of the examples in Section 2 concerned the case of $\mathcal{I} = \mathcal{T}$. To let $\mathcal{I}$ be a proper subset of $\mathcal{T}$ is to ask the algorithm to produce a particularly close individual pairing $(\tau, \gamma)$ for a specific subset of treated subjects $\tau \in \mathcal{I}$. When might this be useful?

5.1 Case Conferences in Medicine and Other Forms of Thick Description

In an effort to audit the cost and quality of surgery at 217 U.S. hospitals, Silber et al. (2014) sampled a representative template of 300 surgical patients, then matched surgical patients at each hospital to the template, producing a $300 \times 217$ array with rows defined by the template patient and columns defined by the hospital. In this way, each hospital could compare its own performance with its own patients to the performance of other hospitals when performing the same surgical procedure on similar patients.

A comparison of 300 patients is useful. However, a standard part of medical quality control is the case-conference, in which the care of one patient is reviewed in detail. The *New England Journal of Medicine* often publishes its "Case Reports from the Massachusetts General Hospital," which are essentially case-conferences in essay form. Template matching offers to enhance a case-conference by permitting a comparison of a case in one's own hospital and very similar cases in other hospitals.

To be useful in such a comparative case-conference, the pair of patients in different hospitals would need to be very close as individuals, not a small part of a large counterbalancing scheme for 300 patients involving propensity scores and fine balance constraints. By letting $\mathcal{I} \subset \mathcal{T}$ be a subset of, say, $|\mathcal{I}| = 10$ interesting template patients, one could ensure exceptionally close individual pairs for all hospitals for these 10 cases, while using propensity scores and fine balance to produce comparable groups of 300 patients.
More generally, thick, narrative, or ethnographic description of a small number of closely matched pairs is often an enlightening counterpoint to quantitative analysis of many pairs. For discussion see Rosenbaum and Silber (2001) and Weller and Barnes (2014).

5.2 Close Pairs When Matching With Multiple Controls

When matching each treated individual to several controls, say three controls, it can be helpful that each matched set contains one control who is very close to the treated subject in terms of specific covariates. For instance, this can be advantageous when studying effect modification involving a few covariates (Hsu, Small, and Rosenbaum 2013; Hsu et al. 2015) or when seeking reduced sensitivity to unmeasured biases by reducing within-pair heterogeneity of response (Rosenbaum 2005). For the other two controls in the same matched set, the covariates would be balanced as a group, as in Table 2, but might not be close as individual pairs. In this way, one has, in a single study, both the larger sample size of a 3-to-1 match and the less heterogenous pairing of a 1-to-1 match.

To create such a match, replace $T = \{t_1, \ldots, t_T\}$ by $T = \{t_1, \ldots, t_T, t'_1, \ldots, t'_T\}$ where $t'_i$ is a second copy of treated subject $t_i$, and let $I = \{t_1, \ldots, t_T\}$ contain the first copies. The network optimization would operate as if $t_i$ and $t'_i$ were distinct individuals, and after matching their controls would be joined to form matched set $t$. For the 3-to-1 match described above, $t_i$ would have demand $d(t_i) = -1$ and $t'_i$ would have demand $d(t'_i) = -2$. Use of Algorithm A with $I = \{t_1, \ldots, t_T\}$ and $h_k$ equal to the robust Mahalanobis distance would minimize the maximum distance for the $t_i \in I$, trying to ensure one very close control for every treated individual in every matched tripel. When matching with variable numbers of controls, the duplicates $t'_i$ demands could be determined by the entire number of Keele, Pimentel, and Yoon (2015).

6. Discussion: Searching to Find the Best Feasible Constraint

Optimal matching in observational studies minimizes the total of within-pair covariate distances subject to various constraints that enforce covariate balance or exact matches. This is done by finding a minimum cost flow in a network, a standard problem in operations research for which good polynomial-time algorithms exist. A match that is optimal in this sense may include a moderate number of pairs with large covariate distances. Here, a method has been proposed and illustrated for adding one or more new constraints that prevent any pairs from exhibiting larger covariate distances than are strictly needed to satisfy the constraints. Stated more precisely, the new constraint forces the match to achieve the minimum possible value of the maximum within-pair covariate distance among all matches that satisfy the constraints; then, with that additional constraint in place, the match minimizes the total within-pair covariate distance. To set this additional constraint, the needed minimax value is determined by a binary search involving repeatedly constructing a matched sample of a simple form; then, with this value determined, edges are removed from the matching network forbidding matches that exceed the minimax value. Removing edges from a matching network improves the worst-case time bound for computing the minimum cost flow. Here, removing edges means forbidding consideration of many matched pairs that we did not want anyway. As discussed at the end of Section 4.4, the algorithm may be accelerated by forcing the maximum distance in the match to exceed the minimax value by at most a small error $\delta$. In the example with 1391 matched pairs, the minimax constraint avoided dozens of large covariate distances and exhibited excellent covariate balance. Various generalizations were discussed and illustrated, including quantile constraints and iterated minimax constraints.

Supplementary Materials

Online supplementary material contains an R package, aamatch, that includes: (i) a limited implementation of the method, (ii) the data used in the example, (iii) an example that reproduces some of the matched comparisons for the example.

Acknowledgments

This work is supported by the U.S. National Science Foundation.

References

Bang-Jensen, J., and Gutin, G. (2009), Digraphs: Theory, Algorithms and Applications (2nd ed.), New York: Springer. [73]
Bertsekas, D. P., and Tseng, P. (1988), “The Relax Codes for Linear Minimum Cost Network Flow Problems,” Annals of Operations Research, 13, 125–190. Fortran and C code: http://www.mit.edu/~dimitrib/home.html. Available in R via the optmatch package. [73]
Burkard, R., Dell’Amico, M., and Martello, S. (2009), Assignment Problems, Philadelphia, PA: SIAM. [71]
Cahuc, P., Carcillo, S., and Zylberberg, A. (2014), Labor Economics, Cambridge, MA: MIT Press. Available at http://www.labor-economics.org/ [69]
Garfinkel, R. S. (1971), “An Improved Algorithm for the Bottleneck Assignment Problem,” Operations Research, 9, 1747–1751. [71,75]
Hansen, B. B. (2007), “Optmatch,” R News, 7, 18–24. (R package optmatch) [66]
Hansen, B. B., and Klopfner, O. S. (2006), “Optimal Full Matching and Related Designs via Network Flows,” Journal of Computational and Graphical Statistics, 15, 609–627. [66,69]
Hsu, J. Y., Small, D. S., and Rosenbaum, P. R. (2013), “Effect Modification and Design Sensitivity in Observational Studies,” Journal of the American Statistical Association, 108, 135–148. [68,77]
Hsu, J. Y., Zubizarreta, J. R., Small, D. S., and Rosenbaum, P. R. (2015), “Strong Control of the Family-Wise Error Rate in Observational Studies That Discover Effect Modification by Exploratory Methods,” Biometrika, 102, 767–782. [68,77]
Kao, M. Y., Lam, T. W., Sung, W. K., and Ting, H. F. (2001), “A Decomposition Theorem for Maximum Weight Bipartite Matchings,” SIAM Journal on Computing, 31, 18–26. [68]
Keele, L., Pimentel, S., and Yoon, F. (2015), “Variable-Ratio Matching With Fine Balance in a Study of Peer Health Exchange,” Statistics in Medicine, 34, 4070–4082. [74,77]
Lalive, R., Van Ours, J., and Zweimüller, J. (2006), “How Changes in Financial Incentives Affect the Duration of Unemployment,” Review of Economics and Statistics, 73, 1009–1038. [69]
Lu, B., Greer, R., Xu, X., and Beck, C. (2011), “Optimal Nonbipartite Matching and its Statistical Applications,” The American Statistician, 65, 21–30. (R package matchit) [66]
Neuman, M. D., Rosenbaum, P. R., Ludwig, J. M., Zubizarreta, J. R., and Silber, J. H. (2014), “Anesthesiology Technique, Mortality, and Length of Stay
Pimentel, S. D., Kelz, R. R., Silber, J. H., and Rosenbaum, P. R. (2015), “Large, Sparse Optimal Matching With Refined Covariate Balance in an Observational Study of the Health Outcomes Produced by New Surgeons,” *Journal of the American Statistical Association*, 110, 515–527. [66,68,73,74]

Rosenbaum, P. R. (1989), “Optimal Matching in Observational Studies,” *Journal of the American Statistical Association*, 84, 1024–1032. [66,69,73]

——— (1991), “A Characterization of Optimal Designs for Observational Studies,” *Journal of the Royal Statistical Society*, Series B, 53, 597–610. [69,73]

——— (2005), “Heterogeneity and Causality,” *American Statistician*, 59, 147–152. [68,77]

——— (2010), *Design of Observational Studies*, New York: Springer. [66,69]

Rosenbaum, P. R., and Silber, J. H. (2001), “Matching and Thick Description in an Observational Study of Mortality After Surgery,” *Biostatistics*, 2, 217–232. [68,77]

Silber, J. H., Rosenbaum, P. R., Clark, A. S., Giantonio, B. J., Ross, R. N., Teng, Y., Wang, M., Niknam, B. A., Ludwig, J. M., Wang, W., Even-Shoshan, O., and Fox, K. R. (2013), “Characteristics Associated With Differences in Survival Among Black and White Women With Breast Cancer,” *Journal of the American Medical Association*, 310, 389–397. [66]

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. [68]

Stuart, E. A. (2010), “Matching Methods for Causal Inference,” *Statistical Science*, 25, 1–21. [66]

Weller, N., and Barnes, J. (2014), *Finding Pathways: Mixed-method Research for Studying Causal Mechanisms*, New York: Cambridge University Press. [68,77]

Yang, D., Small, D. S., Silber, J. H., and Rosenbaum, P. R. (2012), “Optimal Matching With Minimal Deviation From Fine Balance in a Study of Obesity and Surgical Outcomes,” *Biometrics*, 68, 628–636. (Archived R package finebalance) [66,69,74]

Zubizarreta, J. R. (2012), “Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery,” *Journal of the American Statistical Association*, 107, 1360–1371. (R software mipmatch) [66,69,73]