Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery

José R. Zubizarreta

Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery

José R. ZUBIZARRETA

This article presents a new method for optimal matching in observational studies based on mixed integer programming. Unlike widely used matching methods based on network algorithms, which attempt to achieve covariate balance by minimizing the total sum of distances between treated units and matched controls, this new method achieves covariate balance directly, either by minimizing both the total sum of distances and a weighted sum of specific measures of covariate imbalance, or by minimizing the total sum of distances while constraining the measures of imbalance to be less than or equal to certain tolerances. The inclusion of these extra terms in the objective function or the use of these additional constraints explicitly optimizes or constrains the criteria that will be used to evaluate the quality of the match. For example, the method minimizes or constrains differences in univariate moments, such as means, variances, and skewness; differences in multivariate moments, such as correlations between covariates; differences in quantiles; and differences in statistics, such as the Kolmogorov–Smirnov statistic, to minimize the differences in both location and shape of the empirical distributions of the treated units and matched controls. While balancing several of these measures, it is also possible to impose constraints for exact and near-exact matching, and fine and near-fine balance for more than one nominal covariate, whereas network algorithms can finely or near-finely balance only a single nominal covariate. From a practical standpoint, this method eliminates the guesswork involved in current optimal matching methods, and offers a controlled and systematic way of improving covariate balance by focusing the matching efforts on certain measures of covariate imbalance and their corresponding weights or tolerances. A matched case–control study of acute kidney injury after surgery among Medicare patients illustrates these features in detail. A new R package called `mipmatch` implements the method.

KEY WORDS: Covariate balance; Matched sampling; Optimal matching; Propensity Score.

1. INTRODUCTION: MOTIVATING EXAMPLE, REVIEW OF MATCHING, OUTLINE

1.1 Motivating Example: Obesity and Acute Kidney Injuries

Does obesity increase the risk of acute kidney injury (AKI) for patients following surgery? As part of an ongoing case–control study of this question, we matched cases of AKI following colectomy to controls without AKI, using a new matching method based on mixed integer programming. We used data from the U.S. Medicare population, from the Obesity and Surgical Outcomes Study (OBSOS; Silber et al. 2011, 2012).

The OBSOS study was designed to compare obese and nonobese surgical patients in Medicare with respect to outcomes, economic measures, and access to surgical care. OBSOS focused on five types of surgery: colectomy for cancer, colectomy not for cancer, thoracotomy, hip replacement without fracture, and knee replacement. Here, colectomy for cancer is used to illustrate the methodology, but the actual case–control study uses all five types of surgery and the same matching methods were applied five times. Specifically, the group with colectomy for cancer consisted of 108 cases with AKI and 1650 controls without AKI.

In OBSOS, the covariates were obtained from claims and chart abstraction from 47 hospitals in Illinois, New York, and Texas. The covariates used in the matched case–control study were: age, black (1 = black), sex (1 = female), emergency (1 = patient was admitted via emergency room), transfer (1 = patient was admitted via transfer), APACHE (a modification of a widely used acute physiology score), risk (risk-of-death score estimated from Medicare claims from all hospitals in Illinois, New York, and Texas, with the exception of the 47 hospitals participating in the study; Hansen 2008), dysfunction (1 = renal dysfunction comorbidity), failure (1 = renal failure comorbidity), and a key quantitative measure of kidney function, eGFR (the estimated glomerular filtration rate, which describes the preoperative flow rate of filtered fluid through the kidney). The eGFR was calculated using Stevens et al.'s (2006) MDRD4 formula, and it is the most important covariate in the study. This measure is used to detect early kidney damage and to monitor kidney status, and as a result, it is closely related with AKI. For this reason, we needed to adjust for eGFR very precisely. Three other important covariates were the nominal covariates that define the specific surgical procedures (left colectomy, right colectomy, and other colectomy), the 47 hospitals involved in the study, and the interaction of history of renal dysfunction and renal failure.

In the case–control study, we matched (a) exactly for the specific surgical procedure, (b) balancing the marginal distribution of the nominal covariates for the 47 hospitals in the study and for the interaction of renal dysfunction and renal failure, (c) balancing as closely as possible all the means of the other medically important covariates, and (d) balancing the entire empirical distribution of eGFR. This article presents a new matching method for observational studies based on mixed integer programming

that finds such matches and even more difficult ones. As we will see in what follows, the use of mixed integer programming methods in place of commonly used network algorithms substantially enlarges the tools available for matching.

1.2 Review of Matching in Observational Studies

In a randomized experiment, units are assigned to treatment or control by the flip of a fair coin, yielding two groups that are similar in terms of observed and unobserved covariates. As a result, in randomized experiments, there is no systematic confounding on the covariates, and association between the treatment and the outcome implies causation. Accordingly, in randomized experiments, it is relatively easy to obtain unbiased estimates of treatment effects.

However, in a wide variety of settings, it is very expensive, impractical, or unethical to randomly assign units to treatment or control, and instead, we have to rely on observational data. In observational studies, the allocation of treatment is outside our jurisdiction as researchers and we merely observe that some units are treated and that others are controls. In these settings, matching is an attempt to recover the randomized experiment that is "hidden" inside the observational study (Hansen 2004).

The main goal of matching in observational studies is to free the comparisons of the outcome variables in the treated and matched controls groups from the confounding effects of differences in their observed covariates (Cochran, Moses, and Mosteller 1983; Rosenbaum 2002; Stuart 2010). With matching, we attempt to replicate a randomized experiment by comparing treated units to controls that are similar in terms of their observed covariates, as if they were randomly assigned to treatment. For this, we need to make the empirical distributions of the observed covariates of these two groups as similar as possible. Note that differences in unobserved covariates are not addressed by matching, and other methods such as sensitivity analysis need to be used instead (see Rosenbaum 2010, sec 14 for an introduction).

Matching has several attractive features, as discussed by Stuart (2010) and Lu et al. (2011). First, it encourages critical discussion by medical experts who may have difficulty understanding statistical models, but have greater knowledge of the subject matter than do statisticians. Second, it forces researchers to look closely at the data, drawing their attention to issues such as the common support of the covariates and helping them to understand better the constraints imposed by the data on the analysis. Third, since matching is completed before outcomes are examined, it helps to separate the design and analysis stages of the study and thus to prevent inappropriate manipulation of the data. Fourth, as pointed out by Rosenbaum and Rubin (1985), despite the complexity of some matching algorithms, very simple methods can be used to analyze the outcomes, and typical analysis of matched samples do not require the parametric assumptions of most regression methods (Lu et al. 2011). Fifth, unlike other statistical methods of adjustment, matching facilitates the combination of qualitative and quantitative methods because it maintains intact the unit of analysis. Indeed, after matching, it is possible to conduct ethnographic or thick description of cases, and this in-depth investigation in turn can aid statistical analysis (Rosenbaum and Silber 2001). Sixth, using

existing inferential techniques is straightforward if matching is done without replacement (Hansen and Klopfer 2006). Finally, the resulting structure after matching is simple enough to allow for sensitivity analysis to unobserved covariates, which is the major concern in observational studies.

However, in spite of these attractive features, finding a good match that balances the empirical distributions of all the observed covariates in the treated and matched control groups can be very difficult in practice. Current optimal matching methods based on network algorithms pursue covariate balance indirectly, by minimizing the total sum of distances between treated units and matched controls, and hope that this will result in balance. With these methods, covariate balance is improved by iteratively estimating a distance between treated units and potential controls, finding the matches, and checking balance until a satisfactory level is achieved. The problem is that this iterative process usually involves a considerable amount of guesswork, and it is common to see that balancing some covariates worsens the balance of other covariates (Diamond and Sekhon 2005; Hill 2011; Icacus, King, and Porro 2012). In this sense, there is little control over the matching process. Moreover, with this procedure, balancing means is already difficult, but one would like to go beyond that and adjust for nonlinearities and nonadditivities in the matched samples by explicitly balancing other statistics such as the Kolmogorov–Smirnov (K–S) statistics or correlations between covariates. The new matching method presented in this article deals with these issues and offers a controlled and systematic way of improving balance based on several features of the empirical distributions of the observed covariates.

1.3 Outline: Optimal Matching With Controlled Balance for Multiple Features of the Empirical Distributions of the Observed Covariates

Unlike widely used matching methods based on network algorithms, which attempt to achieve covariate balance by minimizing the total sum of distances between treated units and matched controls, this new method achieves covariate balance directly, either by minimizing both the total sum of distances and a weighted sum of specific measures of covariate imbalance, or by minimizing the total sum of distances while constraining the measures of imbalance to be less than or equal to certain tolerances. The inclusion of these extra terms in the objective function or the use of these additional constraints explicitly optimizes or constrains the criteria that will be used to evaluate the quality of the match. For instance, with this method, it is possible to minimize or constrain differences in univariate moments, such as means, variances, and skewness; differences in multivariate moments, such as correlations between covariates; differences in quantiles; and differences in statistics, such as the K–S statistics, to reduce the differences in both location and shape of the empirical distributions of the case and matched controls samples. With the proposed method, it is possible to balance not one but several of the previous criteria simultaneously.

Furthermore, with the proposed method, it is possible to do this while also satisfying constraints for exact and near-exact matching, fine and near-fine balance, for multiple covariates. Roughly speaking, exact matching refers to matching every treated unit to one or more controls with the same value of

a nominal covariate while balancing exactly the marginal distribution of that covariate (see Rosenbaum 2010, sec 7.2 for a discussion on exact matching). When exact matching is not feasible, it might still be desirable to approximate it with near-exact matching (see Rosenbaum 2010, sec 9.2 for an introductory discussion on near-exact matching, and Zubizarreta et al. 2011 for using it to match for several sparse nominal covariates). In a similar way to exact matching, fine balance constrains the marginal distributions of a nominal variable to be the same in the case and control groups, but without constraining who is matched to whom (see Rosenbaum, Ross, and Silber 2007, sec 10; Rosenbaum 2010). When fine balance is not feasible, it is possible to approximate it with near-fine balance (Yang et al. 2012). With the proposed matching method, it is possible to achieve fine or near-fine balance for more than one nominal covariate, whereas network algorithms can finely or near-finely balance only a single nominal variable.

From a practical standpoint, this method offers a controlled and systematic way of improving covariate balance by focusing the matching efforts on certain measures of covariate imbalance and their corresponding weights or tolerances. From a computational point of view, with a current implementation of this method, it is possible to solve large problems, comparable to those solved by network optimization, in a comparable time. This opens a new dimension in combinatorial optimization for matching in observational studies. All these features are implemented in the new `mipmatch` package in R.

Following this introduction, the rest of the article is organized as follows. Section 2 gives some background on mixed integer programming and describes recent developments in computation that have made it possible to solve large instances of these problems. Section 3 presents a general mixed integer program (MIP) for matching. Section 4 discusses several covariate imbalance terms that can be directly targeted with this method and illustrates them using the case–control study of Medicare patients with AKIs. Section 5 analyzes the outcomes from this study. Section 6 extends the method to matching with variable ratio. Section 7 concludes with a summary and a discussion.

2. BACKGROUND ON MIXED INTEGER PROGRAMMING

For several decades, a combinatorial optimization problem, such as optimal matching of cases to controls, was considered tractable if the most difficult problem of a given size could be solved in a number of arithmetic steps that grows as a polynomial in the size of the problem. The minimum total distance matching problem (the assignment problem) and related "network" optimization problems are tractable in this sense, whereas the more general MIP is not. No doubt, for this reason, statisticians used optimal assignment algorithms but not MIPs for matching. The sharp line between tractable and intractable combinatorial problems has been blurred in recent years.

In observational studies, the optimal assignment algorithm matches every case to one or more controls to minimize the total sum of distances of the matched groups (Rosenbaum 1989). While it is possible to solve this problem by means of the simplex algorithm (due to the totally unimodular constraint matrix,

which implies that all its square submatrices have a determinant equal to -1, 0, or 1, and thus, that the problem has integer basic solutions), there exist more specialized algorithms that take advantage of its particular structure. One of them is the Hungarian method proposed by Kuhn (1955), which can be modified to achieve a worst time bound of order $O(C^3)$, where C is the number of potential controls, and which is equivalent to the worst time required to multiply two $C \times C$ matrices in the customary way. An often faster algorithm is the auction algorithm proposed by Bertsekas (1981). Using the auction algorithm, optimal pair matching was made widely available to the statistical community through the `optmatch` package in R by Hansen (2007).

On the other hand, mixed integer programming has been rarely used for matching in observational studies. An exception is Li, Propert, and Rosenbaum (2001), who used penalty optimization in the context of balanced risk set matching. However, in recent years, there have been important advances in linear programming (LP), which is the facilitating technology for solving MIPs (Bixby et al. 2000; Bertsimas and Weismantel 2005). Using LP as a tool, modern solvers integrate branch-and-bound, and cutting plane algorithms to solve general MIPs. The branch-and-bound algorithm iteratively partitions the solution space into smaller MIPs by rounding the solution of LP relaxations, and discards subsets of the candidate solutions by using bounds of the optimized quantity. The cutting plane algorithm iteratively refines the feasible set to find the convex hull of the mixed integer solutions. The logic behind integrating these two algorithms is, on the one hand, to find a better approximation of the convex hull and thus increase the chances that fewer partitions of the solution space are needed with the branch-and-bound algorithm, and, on the other hand, to avoid the slow convergence rates of the addition of too many constraints, which presents numerical difficulties for solvers with the cutting plane algorithm (Bixby and Rothberg 2007; Linderoth and Lodi 2010).

Besides using sophisticated branching and cutting plane generation strategies, the basic components of modern MIP solvers are: presolving (which tightens the original formulation of the problem), heuristics (which try to find good feasible solutions at several stages of the solving process), and parallel implementation (which solves independently several nodes of the branch-and-bound solution tree) (Bixby and Rothberg 2007; Linderoth and Lodi 2010). These advances combined with greater computational power make mixed integer optimization a new tool for matching in observational studies.

3. A MIXED INTEGER PROGRAM FOR OPTIMAL MATCHING: STATEMENT OF THE PROBLEM

Let $\mathcal{T} = \{t_1, \ldots, t_T\}$ be the set of treated units or cases, and $\mathcal{C} = \{c_1, \ldots, c_C\}$, the set of potential controls, with $T \leq C$. Define $\mathcal{P} = \{p_1, \ldots, p_P\}$ as the set of labels of the P observed covariates. Each treated unit $t \in \mathcal{T}$ has a vector of observed covariates $\boldsymbol{x}_t = \{x_{t,p_1}, \ldots, x_{t,p_P}\}$, and each control $c \in \mathcal{C}$ has a similar vector $\boldsymbol{x}_c = \{x_{c,p_1}, \ldots, x_{c,p_P}\}$. Based on these covariates, there is a distance $0 \leq \delta_{t,c} < \infty$ between treated and control units. We want to find an assignment $\alpha : \mathcal{T} \to \mathcal{C}$ that matches each treated unit to a number of different controls, while minimizing or constraining several discrepancy measures

between the empirical distributions of the observed covariates of the treated units and the matched controls.

Unlike the majority of optimal matching methods, which define these discrepancy measures solely as a sum of distances between treated units and matched controls, we express them as both this sum of distances and a number of specific covariate imbalance terms. As we will see, these covariate imbalance terms can enter the matching problem through the objective function to directly minimize imbalances or via several constraints to directly constrain them. This approach allows us to have more control over the covariate balancing process by explicitly optimizing or constraining the criterion that will be used to judge the quality of the match, such as differences in means after matching or differences in the entire empirical distributions of the matched units for several covariates. This allows for a fine-grained adjustment of the covariates in the design of an observational study.

To find the optimal assignment $\alpha^*(\cdot)$, we introduce the decision variable $a_{t,c}$, which is 1 if treated unit t is assigned to control c, and 0 otherwise. We solve the optimization problem

$$
\begin{aligned}
&\underset{a}{\text{minimize}} && \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \delta_{t,c} a_{t,c} + \sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a}) && (1.1) \\
&\text{subject to} && \sum_{c \in \mathcal{C}} a_{t,c} = m, \ t \in \mathcal{T} && (1.2) \\
& && \sum_{t \in \mathcal{T}} a_{t,c} \leq 1, \ c \in \mathcal{C} && (1.3) \\
& && a_{t,c} \in \{0, 1\}, \ t \in \mathcal{T}, c \in \mathcal{C} && (1.4) \\
& && v_j(\boldsymbol{a}) \leq \varepsilon_j, \ j \in \mathcal{J}. && (1.5)
\end{aligned}
\quad (1)
$$

In the objective function, (1.1), $\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \delta_{t,c} a_{t,c}$ is the total sum of distances between treated units and matched controls, and $\sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a})$ is a weighted sum of specific measures of covariate imbalance. In this second term, the $\mu_i(\cdot)$'s are functions that define the measures of covariate imbalance, and the ω_i's are importance weights that each of these measures receive. The argument of each $\mu_i(\cdot)$ is $\boldsymbol{a}$, the vector of all decision variables $a_{t,c}$. The specific form of each $\mu_i(\cdot)$ will vary from problem to problem, depending on what aspects of covariate balance we want to optimize.

In the constraints, (1.2) requires that each treated unit is matched to m controls, and (1.3), that each control is used at most once; (1.4) are integrality constraints. All together, these constraints require matching with a fixed 1:m ratio, but other constraints can be written for matching with a variable ratio (this is discussed in Section 6). In the last set of constraints, (1.5), the $v_j(\cdot)$'s are measures of covariate imbalance analogous to the $\mu_i(\cdot)$'s in (1.1), and the ε_j's are scalar tolerances that define the maximum level of imbalance allowed for each $v_j(\cdot)$. The constraints in (1.5) force each measure of covariate imbalance $v_j(\cdot)$ to be no greater than a given tolerance ε_j, for all $j \in \mathcal{J}$. Note that if both these constraints, (1.5), and the second term of the objective function, $\sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a})$, were omitted, then the above optimization problem would become a version of the optimal assignment problem commonly solved in matching in observational studies. Having these additional constraints and extra terms in the objective function will generally make the above problem a MIP.

With this MIP, we have two ways of targeting covariate balance: through the objective function, by specifying covariate imbalance terms, $\mu_i(\cdot)$, and giving them appropriate weights, ω_i; and via additional constraints, defined by the $v_j(\cdot)$'s and their corresponding tolerances, the ε_j's. Mathematically, it is the same to represent a covariate imbalance term as $\mu_i(\cdot)$ in the objective function or as $v_j(\cdot)$ in the constraints; however, they yield different approaches to achieve balance. In the next section, we discuss both approaches and show how they can be used to achieve different forms of covariate balance.

4. OBJECTIVE FUNCTIONS AND CONSTRAINTS TO ACHIEVE DIFFERENT FORMS OF COVARIATE BALANCE

4.1 Balancing Univariate Moments

As discussed in the review, by matching, we attempt to make the empirical distributions of the observed covariates of the treated units and matched controls as similar as possible. To accomplish this, a first thought is to balance the means of both groups. For this, let $\mathcal{I} \subseteq \mathcal{P}$ denote a subset of the observed covariates and write

$$
\sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a}) = \sum_{i \in \mathcal{I}} \omega_i \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},i} \right|, \quad (2)
$$

where $\overline{x}_{\mathcal{T},i}$ is the mean of covariate i for the treated units. Here, Equation (2) is the weighted sum of the absolute treated-minus-control difference in means for the covariates in $\mathcal{I}$. To linearize (2), we introduce one additional decision variable z_i for each absolute-value term, and solve

$$
\begin{aligned}
&\underset{a,z}{\text{minimize}} && \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \delta_{t,c} a_{t,c} + \sum_{i \in \mathcal{I}} \omega_i z_i \\
&\text{subject to} && \sum_{c \in \mathcal{C}} a_{t,c} = m, \ t \in \mathcal{T} \\
& && \sum_{t \in \mathcal{T}} a_{t,c} \leq 1, \ c \in \mathcal{C} \\
& && z_i \geq \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},i}, \ i \in \mathcal{I} \\
& && z_i \geq - \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i} a_{t,c}}{mT} + \overline{x}_{\mathcal{T},i}, \ i \in \mathcal{I} \\
& && a_{t,c} \in \{0, 1\}, \ t \in \mathcal{T}, c \in \mathcal{C}.
\end{aligned}
\quad (3)
$$

Basically, this transformation forces each z_i to be as close as possible to the largest of the two bounds $\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},i}$ and $-\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i} a_{t,c}}{mT} + \overline{x}_{\mathcal{T},i}$, reducing the difference in means after matching for the covariates in question. This kind of transformation is common in linear optimization, and in our setting, it is used repeatedly to minimize different covariate imbalance terms (see section 1.3 of Bertsimas and Tsitsiklis (1997) for a discussion of linear optimization problems involving absolute values).

We used this formulation in our case–control study to find an optimal pair match that directly balanced the means of all the observed covariates. The results are summarized in Table 1. In this table, the first set of columns presents the means of the cases and controls before matching, showing that the two

Table 1. Balancing the means of the covariates

Covariate	Before matching			Match 1			Match 2			Match 3			Match 4		
	ω_i	$\overline{x}_{\mathcal{T},i}$	$\overline{x}_{\mathcal{C},i}$	ω_i	$\overline{x}_{\mathcal{T},i}$	$\overline{x}_{\mathcal{C},i}$	ω_i	$\overline{x}_{\mathcal{T},i}$	$\overline{x}_{\mathcal{C},i}$	ω_i	$\overline{x}_{\mathcal{T},i}$	$\overline{x}_{\mathcal{C},i}$	ω_i	$\overline{x}_{\mathcal{T},i}$	$\overline{x}_{\mathcal{C},i}$
Age	—	74.19	73.22	0	74.19	74.29	0	74.19	74.18	0	74.19	74.17	10	74.19	74.19
Black	—	0.12	0.08	0	0.12	0.12	0	0.12	0.11	0	0.12	0.12	0	0.12	0.12
Sex	—	0.39	0.47	0	0.39	0.37	0	0.39	0.39	0	0.39	0.36	15	0.39	0.39
Emergency	—	0.32	0.18	0	0.32	0.31	0	0.32	0.31	0	0.32	0.30	15	0.32	0.32
Transfer	—	0.02	0.01	0	0.02	0.02	0	0.02	0.02	0	0.02	0.02	0	0.02	0.02
eGFR	—	54.21	73.60	0	54.21	58.82	10	54.21	54.21	10	54.21	54.21	10	54.21	54.21
APACHE	—	35.73	28.69	0	35.73	33.56	10	35.73	35.73	10	35.73	35.73	10	35.73	35.73
Risk	—	−2.97	−3.82	0	−2.97	−3.08	10	−2.97	−2.97	10	−2.97	−2.97	10	−2.97	−2.97
Dysfunction	—	0.33	0.04	0	0.33	0.26	10	0.33	0.32	10	0.33	0.33	10	0.33	0.33
Failure	—	0.31	0.03	0	0.31	0.21	10	0.31	0.28	20	0.31	0.31	50	0.31	0.31

$\overline{x}_{\mathcal{T},i}$ and $\overline{x}_{\mathcal{C},i}$ are the means of the covariates for the case and control units, respectively. Match 1 uses the assignment algorithm. Match 2 uses MIP to directly balance the means with weights $\boldsymbol{\omega} = (0, 0, 0, 0, 0, 10, 10, 10, 10, 10)$. Matches 3 and 4 also use MIP but improve upon match 2 with weights $\boldsymbol{\omega} = (0, 0, 0, 0, 0, 10, 10, 10, 10, 25)$ and $\boldsymbol{\omega} = (10, 0, 15, 15, 0, 10, 10, 10, 10, 50)$, correspondingly. In match 4, the differences in means are reduced to zero for all the covariates.

groups are quite different in terms of all the covariates, especially eGFR, APACHE, risk, dysfunction, and failure. The second set of columns shows the means of the cases and controls but after a first match obtained by solving the standard assignment problem. We observe an improvement in the balance of most of the covariates; however, the means of eGFR, APACHE, risk, dysfunction, and failure are still quite far apart. We improved the balance of these covariates by solving Equation (3) with appropriate weights for each of the absolute differences in means terms.

Broadly speaking, the weights need to be large enough to minimize imbalances to a satisfactory level. Of course, the ideal values will depend on the problem at hand, but general advice is to divide the distance matrix by its mean value, standardize the covariates, and use weights equal to a 10th of the number of treated units. When this is done, a difference in means equal to 1 standard deviation would have a cost in the objective function equivalent to matching a 10th of the treated units. After matching, if covariate balance is not satisfactory, then we can match again using higher weights, directly targeting those covariates for which balance needs to be improved. In general, smaller imbalances will be at the expense of a higher total sum of distances, but in most applications, covariate balance is the main goal of matching. By using mixed integer programming for matching, one has a tool to regulate this tradeoff.

Following these guidelines, we gave a weight of 10 to each term corresponding to the differences in means of eGFR, APACHE, risk, dysfunction, and failure. The third set of columns presents these results, showing that all the differences in means are considerably lower, and that only renal failure still has an absolute standardized difference in means larger than 0.1. We corrected this remaining imbalance with a third match that increased the weight of the difference in means of failure to 25. In the fourth set of columns, we show the results of this match and note that all the means are now very close.

Mixed integer programming of the form (1) directly optimizes covariate balance, and by doing so, it achieves greater covariate balance in a few steps. In fact, in only three steps, we have reduced all the standardized differences in means to 0.09 or less.

This is satisfactory by common standards; however, by choosing the weights more carefully, we can find an even tighter match and actually reduce all the standardized differences in means to 0.00. The weights used in this fourth match and the resulting means are shown in the fifth set of columns of Table 1. In the spirit of Love (2004), Figure 1 shows the absolute standardized differences in means for these different matches. As we can see, we have obtained maximum bias reduction for all the covariates.

An alternate approach to balancing the means of the covariates is to force the absolute treated-minus-control difference in means to be less than or equal to prespecified levels or tolerances. For this, let $\mathcal{J}$ denote a subset of the observed covariates, and write

$$\left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,j} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},j} \right| \le \varepsilon_j, \quad j \in \mathcal{J}. \qquad (4)$$

In a similar way to Equation (2), we can linearize these constraints by requiring each of the differences in means $\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,j} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},j}$ and $\overline{x}_{\mathcal{T},j} - \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,j} a_{t,c}}{mT}$ to be less or equal than the tolerance ε_j for all $j \in \mathcal{J}$. In our case–control study, by letting $\varepsilon_j = 0.04$ for all $j \in \mathcal{J}$, where $\mathcal{J}$ is the set of labels of all the covariates, we can force all the standardized differences in means to be approximately zero.

Using additional constraints and choosing tolerances has advantages and disadvantages relative to including extra terms in the objective function and picking weights. The constraint approach has the benefit of interpretability: in general, it is more meaningful to define a maximum allowable level of imbalance than a weight. Also, the resulting problem is simpler in principle: it is a pure integer programming problem, whereas the objective function approach yields a mixed integer one. However, the constraint approach has the drawback that specifying very small tolerances can yield an infeasible problem, whereas with the objective function approach, for any choice of weights, there should be a feasible solution. In any case, the two approaches are not mutually exclusive, as one can use extra terms and weights in the objective function to balance one set of
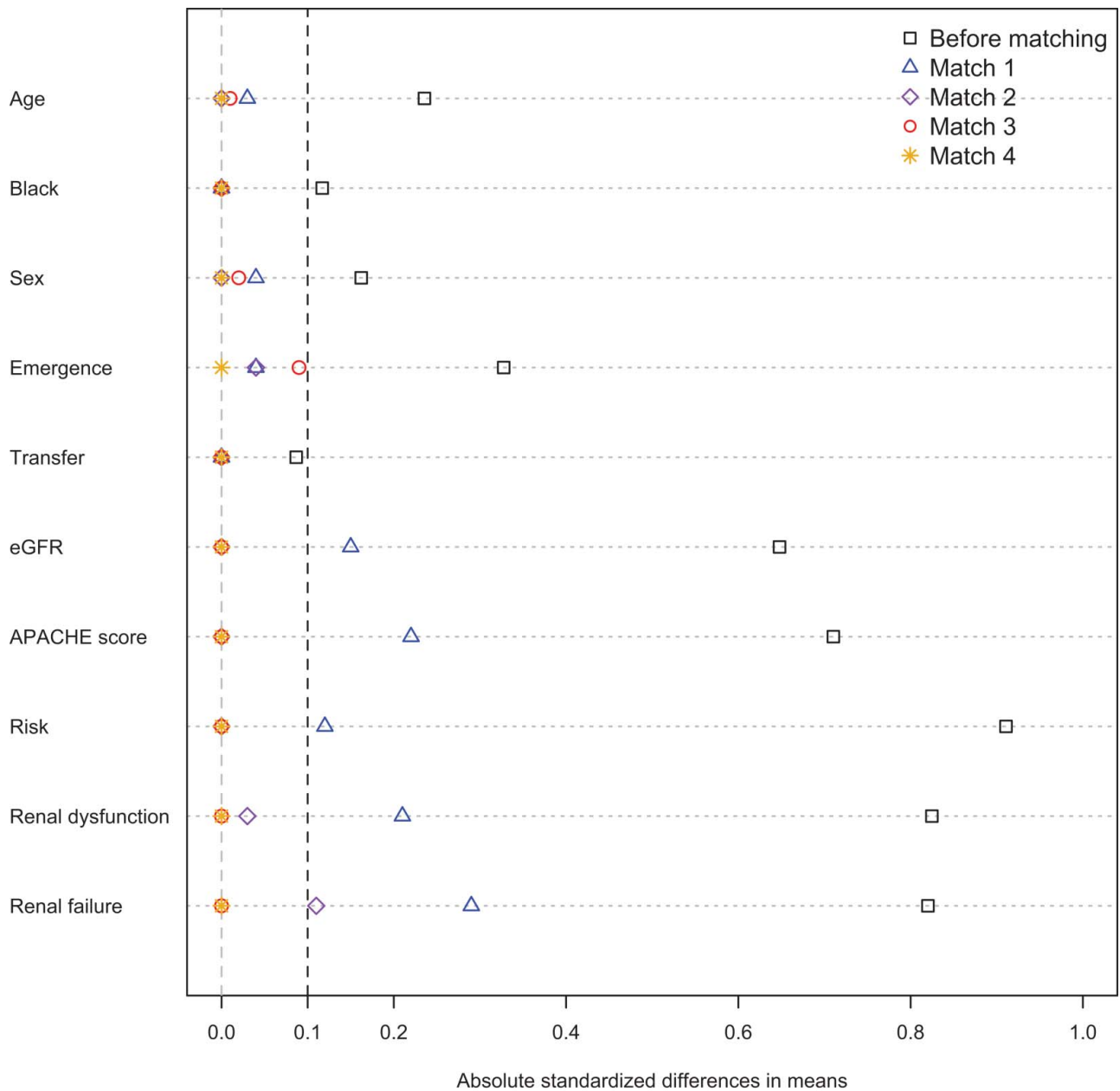
Figure 1. Balancing the means of the observed covariates. Match 1 uses the assignment algorithm. Match 2 uses MIP to directly balance the means with weights $\boldsymbol{\omega} = (0, 0, 0, 0, 0, 10, 10, 10, 10, 10)$. Matches 3 and 4 also use MIP improving upon match 2 with weights $\boldsymbol{\omega} = (0, 0, 0, 0, 0, 10, 10, 10, 10, 25)$ and $\boldsymbol{\omega} = (10, 0, 15, 15, 0, 10, 10, 10, 10, 50)$, respectively. In match 4, the differences in means of all the covariates are reduced to zero. This maximum bias reduction can also be obtained using MIP and forcing the mean imbalances to less or equal than very small tolerances. The online version of this figure is in color.

measures, and at the same time, impose additional constraints with tolerances to force balance on another set. Besides, from a practical standpoint, with very difficult matches, it might be useful to start by finding the matches with weights and then switching to tolerances or vice versa. Since the mathematical expressions of the covariates imbalance terms are similar with both approaches, in what follows for each covariate balance form, we are going to discuss only one of them.

Generalizing our approach for balancing the means through the objective function, we can minimize or constrain the dif-

ferences in the rth moment of a subset of covariates $\mathcal{I}$ if we let

$$\sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a}) = \sum_{i \in \mathcal{I}} \omega_i \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,i}^r a_{t,c}}{mT} - \overline{x}^r_{\mathcal{T},i} \right|, \qquad (5)$$

where $\overline{x}^r_{\mathcal{T},i}$ is the rth moment of covariate i for the treated units, and proceed analogously. Like this, with the proposed MIP formulation for matching, we can simultaneously minimize the differences in means, variances, and skewness.

4.2 Balancing Multivariate Moments

We can minimize the differences in correlations of two covariates, p_1 and p_2, by letting

$$
\begin{aligned}
\sum_{i \in \mathcal{I}} \mu_i(\boldsymbol{a}) = {} & \omega_1 \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,p_1} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},p_1} \right| \\
& + \omega_2 \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,p_1}^2 a_{t,c}}{mT} - \overline{x^2}_{\mathcal{T},p_1} \right| \\
& + \omega_3 \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,p_2} a_{t,c}}{mT} - \overline{x}_{\mathcal{T},p_2} \right| \\
& + \omega_4 \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{x_{c,p_2}^2 a_{t,c}}{mT} - \overline{x^2}_{\mathcal{T},p_2} \right| \\
& + \omega_5 \left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{(x_{c,p_1} \cdot x_{c,p_2}) a_{t,c}}{mT} - \overline{(x_{\cdot,p_1} \cdot x_{\cdot,p_2})}_{\mathcal{T}} \right|,
\end{aligned}
\tag{6}
$$

where $\overline{x^2}_{\mathcal{T},p_1}$ and $\overline{x^2}_{\mathcal{T},p_2}$ are the second moments of covariates p_1 and p_2 for the treated units, respectively, and $\overline{(x_{\cdot,p_1} \cdot x_{\cdot,p_2})}_{\mathcal{T}}$ is the mean of the cross product of covariates p_1 and p_2, also for the treated units. From here, we can proceed as in Section 4.1, linearizing the resulting objective function. The extension to other multivariate moments possibly involving more covariates is analogous.

We used this formulation in our case–control study to minimize the differences in correlations of age and emergency admission. In the controls, before matching, this correlation is equal to 0.037, whereas in the cases, it is equal to -0.068. Our goal here was to find a match such that the correlation between age and emergency admission is as close as possible to -0.068. In a match that directly balanced the means of all the covariates, this correlation resulted in a value of 0.029. However, if in addition to the terms for the mean differences, we use the previous terms for minimizing differences in correlations, with weights equal to (10, 10, 10, 10, 25), the resulting value is exactly -0.068. In this way, we adjusted perfectly for the correlation of the two covariates.

4.3 Balancing Quantiles and Kolmogorov–Smirnov Statistics

It is possible to explicitly balance the quantiles of the distributions of the treated units and matched controls by defining as auxiliary covariates the indicators of the quantiles of the treated units and then minimizing their differences in means as in Section 4.1.

Similarly, we can also minimize a coarse version of the K–S statistics between the empirical distributions of the treated units and matched controls for covariate p if we write

$$
\mu_i(\boldsymbol{a}) = \omega_i \sup_{g_p \in \mathcal{G}_p} \left| F_{\mathcal{T},p}(g_p) - F_{\mathcal{C},p}(g_p, \boldsymbol{a}) \right|,
\tag{7}
$$

where $F_{\mathcal{T},p}(\cdot)$ and $F_{\mathcal{C},p}(\cdot)$ are the empirical cumulative distribution functions of the treated units and matched controls, respectively, for covariate p, and $\mathcal{G}_p = \{g_1, \dots, g_{G_p}\}$ is a grid of G_p quantile values in the treated group for covariate p that defines the cumulative proportions $\mathcal{H}_p = \{h_1, \dots, h_{G_p}\}$. For instance, if $\mathcal{G}_p$ are the deciles of the treated units for covariate p, then

$\mathcal{H}_p = \{0.1, 0.2, \dots, 1\}$. Using Equation (7), the problem can be simplified to one of minimizing the difference in means after matching by defining as auxiliary covariates the indicators of the quantiles defined by $\mathcal{G}_p$ plus an auxiliary decision variable z_i. To minimize the K–S statistics, z_i has to be larger than the differences in means for each of the auxiliary covariates for the quantiles; that is,

$$
\begin{aligned}
\omega_i \mu_i(\boldsymbol{a}) &= \omega_i \sup_{g_p \in \mathcal{G}_p} \left| F_{\mathcal{T},p}(g_p) - F_{\mathcal{C},p}(g_p, \boldsymbol{a}) \right| \\
&= \omega_i z_i \\
&\geq \left| h_{g_p} - \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{\mathbb{1}_{\{x_{c,p} < g_p\}} a_{t,c}}{mT} \right| \quad \forall g_p \in \mathcal{G}_p.
\end{aligned}
\tag{8}
$$

This condition is equivalent to the constraints $\omega_i z_i \geq h_{g_p} - \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{\mathbb{1}_{\{x_{c,p} < g_p\}} a_{t,c}}{mT}$ and $\omega_i z_i \geq -h_{g_p} + \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{\mathbb{1}_{\{x_{c,p} < g_p\}} a_{t,c}}{mT}$, for all $g_p \in \mathcal{G}_p$. A very similar transformation can be applied if the covariate imbalance term for the K–S statistic, $\sup_{g_p \in \mathcal{G}_p} \left| F_{\mathcal{T},p}(g_p) - F_{\mathcal{C},p}(g_p, \boldsymbol{a}) \right|$, is included in the constraints of our matching problem instead of the objective function.

We used this approach in the case–control study to balance the entire distribution of eGFR. For each decile of the distribution of eGFR for the cases, we constrained the proportion of controls from different deciles to be less or equal than 0.01. The results are summarized in Figure 2, which shows the empirical cumulative distributions functions of eGFR for the cases, all the potential controls, and three different matches. The match in the upper-right plot (match 1) was obtained through the classical assignment algorithm, the match in the lower-left plot (match 4) used mixed integer programming to constrain the differences in means of all the covariates to zero, and the match in the lower lower-right plot (match 5) extended this last MIP to constrain both the differences in means of all the covariates and the K–S statistic for eGFR. We observe that the empirical cumulative distributions functions of eGFR for the cases and all the potential controls are quite far apart before matching, but that match 1 finds a subset of controls for which the corresponding distribution is much closer. Balancing all the differences in means as in match 4 improves the lower part of the distribution function, but there seems to be room for improvement in the middle and upper parts. In fact, for match 5, the distribution function of eGFR for the matched controls is almost on top of the corresponding distribution of the cases.

It is worth mentioning that in other observational studies, this continuous covariate can be the propensity score. This formulation makes it possible to capitalize the balancing property of propensity scores, which says that if there is imbalance on the observed covariates, then there is also imbalance on the propensity score (Rosenbaum and Rubin 1983; Stuart 2010). Note that in the current example, there are no propensity scores because it is a case–control study (Joffe and Rosenbaum 1999).

4.4 Balancing Constraints: Exact and Near-Exact Matching; Fine and Near-Fine Balance

In the original problem, in addition to the network constraints, it is possible to include other constraints for exact and near-exact matching, and fine and near-fine balance. Let $x_{\cdot,p}$ be a nominal
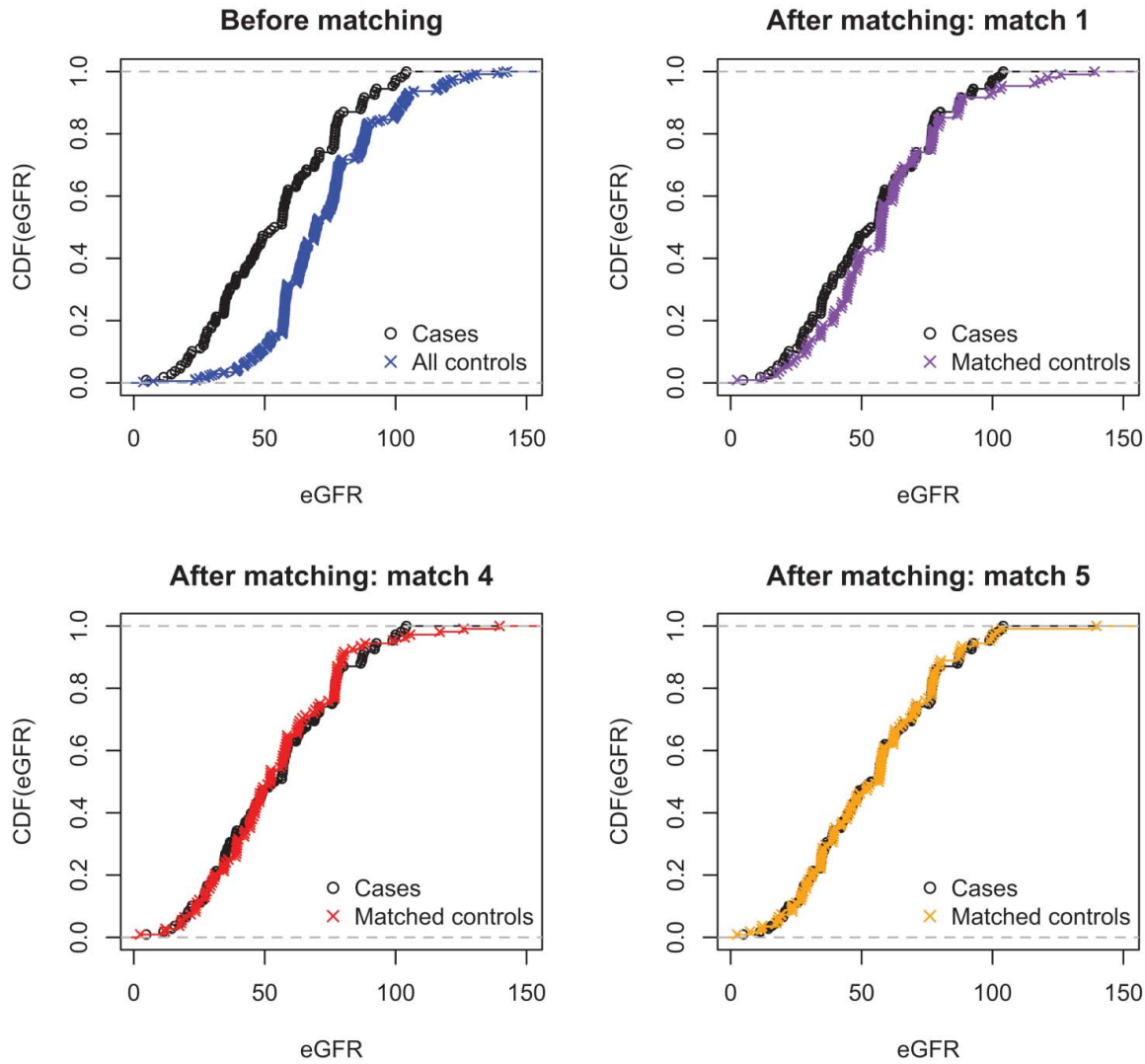
Figure 2. Balancing the K–S statistics for eGFR. Match 1 uses the assignment algorithm. Match 4 uses MIP to directly balance the means of all the covariates through the constraints with tolerances. Match 5 extends match 4 to balance both the means of all the covariates and the K–S statistic for eGFR using the constraint approach. The online version of this figure is in color.

covariate taking integer values $b \in \mathcal{B}$. To match exactly for $x_{\cdot, p}$, we can add the constraints

$$\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \mathbb{1}_{\{x_{t,p}=b\} \bigcap \{x_{c,p}=b\}} = m \sum_{t \in \mathcal{T}} \mathbb{1}_{\{x_{t,p}=b\}} \quad \forall b \in \mathcal{B}, \quad (9)$$

which say that every treated unit has to be matched to m controls with the same values for covariate $x_{\cdot, p}$.

Matching exactly for all the values of $x_{\cdot, p}$ is the ideal thing to do, but this is not always possible. Indeed, for a given b, there might be more treated units than potential controls, making Equation (9) infeasible. To deal with these cases, we can use near-exact matching, which can be defined as the optimal matching obtained by adding the constraints

$$\left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \mathbb{1}_{\{x_{t,p}=b\} \bigcap \{x_{c,p}=b\}} - m \sum_{t \in \mathcal{T}} \mathbb{1}_{\{x_{t,p}=b\}} \right| \le \xi \quad \forall b \in \mathcal{B}, \quad (10)$$

where ξ is the maximum deviation allowed from exact matching for covariate $x_{\cdot, p}$. For example, if $\xi = 2$, we allow the total

number of matched controls to be at most two units away from the total number of treated units with value b.

Note that exact matching is typically done by dividing the data-set into $|\mathcal{B}|$ mutually exclusive and exhaustive parts, obtaining the matches for each of these parts, and then combining these matches. Here, instead, we are matching exactly by solving a single optimization problem. Computationally, this can be very expensive, so a better alternative might be to divide the data-set and proceed as mentioned. On the other hand, it is not possible to do near-exact matching by solving separate problems because the optimal deviation from exact balance is not known in advance. As already seen, our MIP offers a straightforward method to do this.

In the spirit of Rosenbaum, Ross, and Silber (2007) and Yang et al. (2012), it is also possible to match with fine balance and near-fine balance by adding the constraints

$$\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \mathbb{1}_{\{x_{c,p}=b\}} = m \sum_{t \in \mathcal{T}} \mathbb{1}_{\{x_{t,p}=b\}} \quad \forall b \in \mathcal{B} \quad (11)$$

and

$$\left| \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \mathbb{1}_{\{x_{c,p}=b\}} - m \sum_{t \in \mathcal{T}} \mathbb{1}_{\{x_{t,p}=b\}} \right| \leq \xi \qquad \forall b \in \mathcal{B}, \quad (12)$$

respectively.

Following Yang et al. (2012), fine and near-fine balance may be used to balance a nominal variable with many levels, a rare binary variable, or the interaction of several nominal variables, all of which are difficult to control using a distance. With current matching methods, fine balance is achieved by augmenting the distance matrix. Basically, for every control that needs to be removed to achieve fine balance, one extra row with a specific pattern of values equal to zero or infinity is added (see Rosenbaum 2010, sec 16 for details). However, adding extra rows allows for fine balance for one covariate only, or for the interaction of more than one of them, but not for each of them separately, which is considerably more restrictive. On the other hand, with the proposed method, since we are adding one constraint for every nominal covariate that we want to fine balance, we can achieve fine balance for more than one covariate. For the first time, this permits to balance perfectly the marginal distributions of several nominal covariates.

We use these constraints in our case–control study to match exactly for the specific surgical procedures, and with fine balance for both the 47 hospitals in the study and the interaction of renal dysfunction and renal failure. We use these constraints while minimizing the differences in means and in the K–S statistics for eGFR as in match 5. Table 2 shows that the marginal distributions of the surgical procedures are perfectly balanced, while cases are matched to controls within the same surgical procedure. In Table 3, we observe that the 47 hospitals are perfectly balanced, but we are not constraining a case from a given hospital to be matched to a control from the same hospital. Table 4 shows a similar pattern for the interaction of renal dysfunction and renal failure.

4.5 Using a Specific Subset of the Controls

As discussed by Yang et al. (2012), in some cases, it might be desirable to use all the controls with certain covariate values, especially if these controls are close to the treated units in terms of the covariates and are relatively scarce in the control sample. We can force a particular subset of controls to be used if, in addition to the network constraints, we include

$$\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \mathbb{1}_{\{c \in \mathcal{S}\}} = |\mathcal{S}|, \qquad (13)$$

where $\mathcal{S}$ is the subset of indexes of controls to be used.

4.6 Putting It All Together

Under the general framework of Equation (1), all the previous features can be combined and used easily with the new

Table 2. Exact matching for the specific surgery procedure

Procedure	Cases	Controls
Left colectomy	32	32
Other colectomy	14	14
Right colectomy	62	62

Table 3. Fine balance for hospitals

Hospital ID	Cases	Controls	Hospital ID	Cases	Controls
1	3	3	24	3	3
2	3	3	25	5	5
3	3	3	26	0	0
4	0	0	27	1	1
5	4	4	28	0	0
6	1	1	29	3	3
7	0	0	30	1	1
8	0	0	31	3	3
9	3	3	32	3	3
10	1	1	33	1	1
11	3	3	34	1	1
12	1	1	35	5	5
13	0	0	36	10	10
14	7	7	37	4	4
15	3	3	38	2	2
16	1	1	39	1	1
17	3	3	40	1	1
18	4	4	41	2	2
19	4	4	42	1	1
20	5	5	43	1	1
21	1	1	44	2	2
22	4	4	45	0	0
23	4	4	47	0	0

`mipmatch` package in R. This package makes use of IBM's optimization software CPLEX through its R interface `Rcplex` (Corrada 2011). CPLEX is free for academics and students and can be downloaded from the IBM Academic Initiative (*https://www.ibm.com/developerworks/university/academicinitiative/*). `mipmatch` can be downloaded from *www-stat.wharton.upenn.edu/josezubi/*, along with the documentation that describes the steps for installing CPLEX and calling it from R.

All the matches in this article were obtained using `mipmatch` on a laptop with a 2.53-GHz Inter Core 2 Duo processor and a Mac OS X v10.6 Snow Leopard operating system. The final match, which (a) matched exactly for surgical procedure, (b) finely balanced the marginal distributions for the nominal covariates for the 47 hospitals in the study and for the interaction of renal dysfunction and renal failure, (c) explicitly balanced the means of the 10 covariates, and (d) balanced the entire empirical distribution of eGFR, took only 1 min to be solved. As was said in the Introduction, this sample consists of 108 cases and 1650 controls, and corresponds to the patients that had colectomy for cancer in the OBSOS study. If one extended this sample to the patients in OBSOS who underwent any type of colectomy or any type of surgery procedure, the sample sizes would increase

Table 4. Fine balance for the interaction of renal dysfunction and renal failure

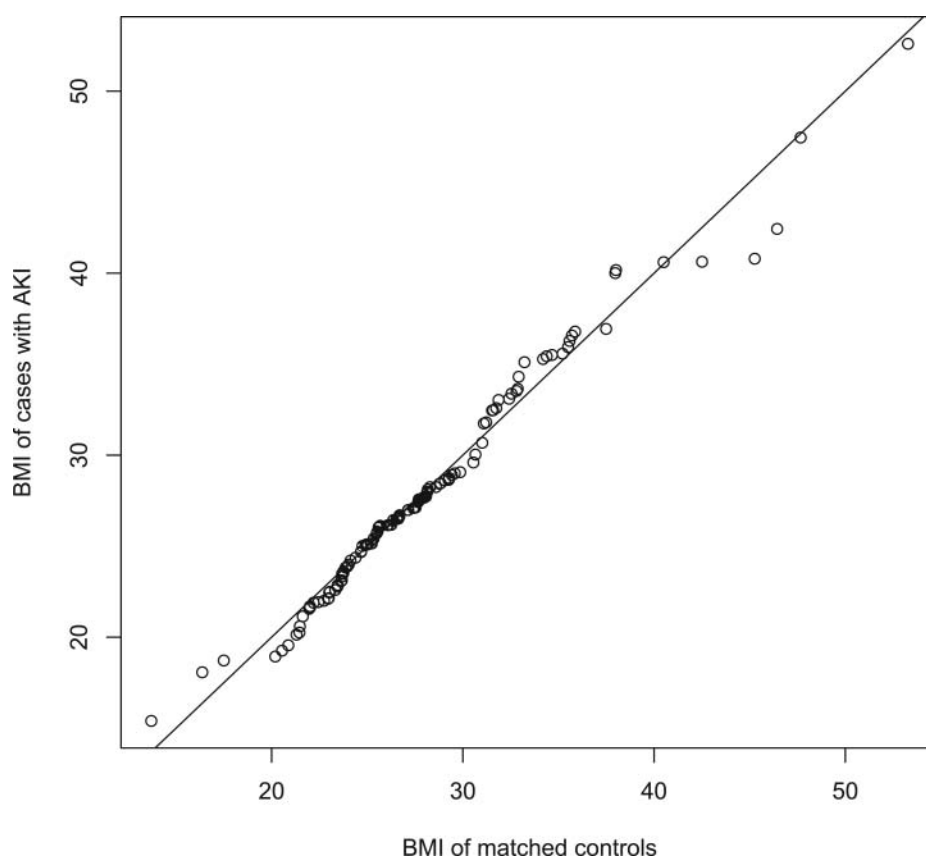| Renal commorbidity	Cases	Controls
Dysfunction = no, failure = no	63	63
Dysfunction = yes, failure = no	11	11
Dysfunction = no, failure = yes	9	9
Dysfunction = yes, failure = yes	25	25

Figure 3. Quantlie–quantile (qq)plot of BMI for cases and matched controls.

to 182 cases and 2437 controls, and 283 cases and 4356 controls, correspondingly. In these problems, the running times for similar matches were 12 and 18 min, respectively.

Of course, computing times in general will depend on the data at hand and the form of covariate balance that one wants to achieve. Nonetheless, general advice to speed up the matching is to split the problem into smaller matching problems by categories of the covariates for exact matching. Since exact matching forces units to be matched within the same categories of the covariates, solving a single problem or splitting them into several smaller ones and then putting them together is almost equivalent. For example, in the previous match that took 18 min, we split into three smaller problems by the categories of surgery taking 10 sec, 1 min, and 1 min each.

Finally, it is worth noting that in all these examples, the ratio of potential controls to cases was quite high, ranging approximately between 13:1 and 17:1. In many problems in practice, however, this ratio is lower and there is a clear tradeoff between balancing one set of covariates and others. In such problems, `mipmatch` provides a useful tool to regulate that tradeoff by directly targeting covariate by choosing weights or tolerances and finding a satisfactory match.

5. RESULTS

In this section, we take a brief look at potential risk factors for AKI following colectomy for cancer. Our central question was whether obese patients are at greater risk of AKI. The results suggest that they are not. Figure 3 is a quantile–quantile

plot of body mass index (BMI) for cases and matched controls. If the distributions of BMIs were equal for cases with AKI and controls, then the points would lie on the $45°$ line. In the figure, there seems to be a departure from this line for BMIs with values smaller than 21, and for higher BMIs with values between 32 and 36, but this association is not significant. In fact, for Wilcoxon's signed rank test with continuity correction, the point estimate is -0.18, with 95% confidence interval of $[-1.45$ to $1.80]$. (For the three categories of obesity, given by BMI $\geq$ 35, $18 \leq$ BMI < 35, and BMI < 18, McNemar's generalized p-value is equal to 0.34; see Fleiss, Levin, and Paik 2003, sec 13, for a presentation of McNemar's test.) On the BMI scale, the previous confidence interval is very narrow around zero: one unit of BMI is approximately 5% of the weight of a person. We also tested for differences in dispersion, finding that they are not significant. In this way, there is no sign that BMI predicts AKI independently of other risk factors. After matching, it is fairly easy to see this result in Figure 3, whereas with a model, it is less easy to see this relation so clearly. Therefore, from a medical standpoint, our results show that it is better to characterize the risk of AKI in terms of factors such as the APACHE score, eGFR, and the histories of renal problems, because after we adjust for them, BMI is irrelevant.

Other risk factors of interest in the study were hypotension and sepsis complications. For hypotension, McNemar's test yields a p-value equal to 0.00004, whereas for sepsis complication, it is equal to 0.0003. Thus, it is important to account for hypotension and sepsis complications in characterizing the risk of AKI following colectomy for cancer. Further analysis

will be conducted in the future to determine the association between other risk factors and AKI after other surgical procedures, such as colectomy without cancer, thoracotomy, hip replacement without fracture, and knee replacement.

6. USING A VARIABLE MATCHING RATIO

The MIP presented in Section 3 can be extended to accommodate other matching structures such as matching with a variable ratio. For this, we use the *entire number* introduced by Yoon (2009), which is the optimal number of controls that should be matched to a treated unit. Extending the notation of (1), we denote the entire number for treated unit t as m_t and calculate it as $m_t = \frac{1-e(x_t)}{e(x_t)}$, where $e(x_t)$ is the propensity score for treated unit t. The entire number has an intuitive interpretation: if $e(x_t) = 1/3$, then treated unit t should be matched to $m_t = \frac{1-1/3}{1/3} = 2$ controls.

Using the entire number, we can modify the objective function and the constraints of our MIP to match with a variable ratio. In the objective function, we minimize $\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \frac{\delta_{t,c}}{m_t} a_{t,c} + \sum_{i \in \mathcal{I}} \omega_i \mu_i(\boldsymbol{a}, m_t)$, where the distances are now normalized by the entire number and the covariate imbalance terms now are also a function of m_t (in a similar way to the distances, the covariate summary measures of the treated units are also normalized by m_t instead of the fixed number m). In the constraints, we replace (1.2) by $\sum_{c \in \mathcal{C}} a_{t,c} = m_t$, $t \in \mathcal{T}$, and modify (1.5) to have different balancing constraints for each stratum of m_t. In this way, following Yoon (2009), within each stratum of m_t, we would match with a fixed $1:m_t$ ratio and, say, use fine balance within each stratum to produce a weighted version of fine balance. In the analysis, we would take each treated unit t minus the average of its m_t controls for each $t \in \mathcal{T}$.

7. SUMMARY AND DISCUSSION

For many years, mixed integer programming problems have been regarded as intractable and impractical problems in statistics. However, as we have seen in this article, this has changed with modern computers and powerful solvers. Although from the viewpoint of computational complexity, general mixed integer programming problems are still labeled as intractable, they have become practical to solve in typical observational studies, such as the case–control study of kidney failure after surgery. As a matter of fact, we can solve matching problems of comparable size in comparable time to those solved by network optimization algorithms, but with several advantages from the perspective of balancing the covariates.

As we have seen, by using mixed integer programming, we can directly balance several features of the marginal and joint distributions of the treated units and matched controls, whereas with network algorithms, we hope to attain this indirectly by minimizing the total sum of distances between units. One way of trying to balance these features without minimizing distances is using propensity scores, but they produce stochastic balance (such as randomization) rather than forcing precise balance; in particular, stochastic balance tends to be inadequate with nominal covariates that have many categories (Zubizarreta et al. 2011). Further, propensity scores do not exist in case–control studies like the one discussed in this article. The proposed method allows us to explicitly optimize the criteria that will

be used to evaluate the quality of the match, and it is possible to do this both in observational studies of treatments and in case–control studies.

As we have discussed, we can use mixed integer programming to directly constrain or minimize differences in univariate moments, such as means, variances, and skewness, and achieve maximum bias reduction on the observed covariates. In addition, we can make the entire distributions of the covariates close by constraining or minimizing differences quantiles or K–S statistics. Furthermore, we can use mixed integer programming to directly balance features of joint distributions by constraining or minimizing differences in multivariate moments, such as correlations between covariates.

While balancing several of these measures, it is also possible to impose constraints for exact and near-exact matching, and fine and near-fine balance for more than one nominal covariate, whereas network algorithms can finely or near-finely balance only a single nominal variable.

Each of these features is new in the context of optimal matching, but they can be combined and pursued simultaneously by solving a single mixed integer programming problem, as implemented in the new R package `mipmatch`.

From a statistical perspective, one of the benefits of using mixed integer programming for matching is related to model-based adjustments, which are commonly used in observational studies of treatments and case–control studies. Rubin (1979) justified matching as a method that can be usefully combined with model-based adjustments to estimate treatment effects. Specifically, Rubin (1979) found that model-based adjustments of matched pair differences are robust to model misspecification. If the matched samples are more closely balanced, the estimates tend to be more robust to model misspecification. Also, as discussed in Snedecor and Cochran (1980, p. 368), when a covariance adjustment model is used for the outcome, the estimate is more precise when matches are more closely balanced. Along these lines, Greevy et al. (2004) showed that covariance adjustment is more efficient when the covariates are more nearly balanced.

Another statistical benefit of using mixed integer programming for matching is that it allows us to emphasize in the matched pairs those covariates that are of greater importance, and as a result, make statistical tests less conservative. For example, based on analysis from separate data-sets, it might be possible to determine whether some covariates produce higher Spearman correlation between the outcomes of matched units. The results in Hollander, Pledger, and Lin (1974) and Lam and Longnecker (1983) suggest that matching the treated and controls units closely on these covariates tends to make statistical tests less conservative. By using mixed integer programming for matching, it is possible to achieve this by emphasizing those covariates that are highly predictive of the outcome in the matched pairs (or matched groups), while also balancing several covariates on aggregate.

Abadie and Imbens (2006) studied asymptotic properties of nearest-neighbor matching estimators commonly used in economics in which the same control is used as a neighbor of multiple treated units, finding that control for many continuous covariates is difficult when nearest-neighbor matching with replacement is used. In contrast, matching without replacement

using mixed integer programming can balance many continuous covariates using the K–S distance, while permitting close matches on some key covariates, perhaps the propensity score; see Rosenbaum and Rubin (1983) and Dehejia and Wahba (1999) for a discussion of propensity scores.

From a practical standpoint, this method eliminates the guesswork involved in current matching methods and offers a controlled and systematic way of improving covariate balance by focusing the matching efforts on certain covariates, measures of imbalance, and corresponding tolerances or weights.

[Received June 2011. Revised March 2012.]

REFERENCES

Abadie, A., and Imbens, G. (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. [1370]

Bertsekas, D. P. (1981), "A New Algorithm for the Assignment Problem," *Mathematical Programming*, 21, 152–171. [1362]

Bertsimas, D., and Tsitsiklis, J. N. (1997), *Introduction to Linear Optimization*, Belmont, MA: Athena Scientific & Dynamic Ideas. [1363]

Bertsimas, D., and Weismantel, R. (2005), *Optimization Over Integers*, Belmont, MA: Athena Scientific & Dynamic Ideas. [1362]

Bixby, R. E., Fenelon, M., Gu, Z., Rothberg, E., and Wunderling, R. (2000), "MIP: Theory and Practice—Closing the Gap," in *System Modelling and Optimization*, eds. M. J. D. Powell, and S. Scholtes, Dordrecht: Kluwer, pp. 19–49. [1362]

Bixby, R. E., and Rothberg, E. (2007), "Progress in Computational Mixed Integer Programming—A Look Back From the Other Side of the Tipping Point," *Annals of Operations Research*, 149, 37–41. [1362]

Cochran, W. G., Moses, L. E., and Mosteller, F. (1983), *Planning and Analysis of Observational Studies*, New York: Wiley. [1361]

Corrada, H. (2011), "Rcplex: R Interface to CPLEX, Package Version 0.2–3 (Contributions From S. Theussl and K. Hornik)," Available at *http://cran.r-project.org/src/contrib/Archive/Rcplex/* [1368]

Dehejia, R., and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [1371]

Diamond, A., and Sekhon, J. S. (2005), "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," Working Paper, *http://sekhon.berkeley.edu/papers/GenMatch.pdf* [1361]

Fleiss, J. L., Levin, B., and Paik, M. C. (2003), *Statistical Methods for Rates and Proportions*, New York: Wiley. [1369]

Greevy, R., Lu, B., Silber, J. H., and Rosenbaum, P. R. (2004), "Optimal Multivariate Matching Before Randomization," *Biostatistics*, 5, 263–275. [1370]

Hansen, B. B. (2004), "Full Matching in an Observational Study of Coaching for the SAT," *Journal of the American Statistical Association*, 99, 609–618. [1361]

——— (2007), "Flexible, Optimal Matching for Observational Studies," *R News*, 7, 18–24. [1362]

——— (2008), "The Prognostic Analogue of the Propensity Score," *Biometrika*, 95, 481–488. [1360]

Hansen, B. B., and Klopfer, S. O. (2006), "Optimal Full Matching and Related Designs Via Network Flows," *Journal of Computational and Graphical Statistics*, 15, 609–627. [1361]

Hill, J. L. (2011), "Bayesian Nonparametric Modeling for Causal Inference," *Journal of Computational and Graphical Statistics*, 20, 217–240. [1361]

Hollander, M., Pledger, G., and Lin, P. (1974), "Robustness of the Wilcoxon Test to a Certain Dependency Between Samples," *The Annals of Statistics*, 2, 177–181. [1370]

Iacus, S. M., King, G. K., and Porro, G. (2012), "Causal Inference Without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 20, 1–24. [1361]

Joffe, M., and Rosenbaum, P. (1999), "Propensity Scores," *American Journal of Epidemiology*, 150, 327–333. [1366]

Kuhn, H. (1955), "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, 2, 83–97. [1362]

Lam, F. C., and Longnecker, M. T. (1983), "A Modified Wilcoxon Rank Sum Test for Paired Data," *Biometrika*, 70, 510–513. [1370]

Li, Y. P., Propert, J. K., and Rosenbaum, P. R. (2001), "Balanced Risk Set Matching," *Journal of the American Statistical Association*, 96, 870–882. [1362]

Linderoth, J. T., and Lodi, A. (2010), "MILP Software," in *Wiley Encyclopedia of Operations Research and Management Science*, eds. J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, New York: Wiley, pp. 3239–3248. [1362]

Love, T. (2004), "Graphical Display of Covariate Balance," Presentation, See *http://chrp.org/love/JSM2004RoundTableHandout.pdf* [1364]

Lu, B., Greevy, R., Xu, X., and Beck, C. (2011), "Optimal Nonbipartite Matching and Its Statistical Applications," *The American Statistician*, 65, 21–30. [1361]

Rosenbaum, P. R. (1989), "Optimal Matching for Observational Studies," *Journal of the American Statistical Association*, 84, 1024–1032. [1362]

——— (2002), *Observational Studies*, New York: Springer. [1361]

——— (2010), *Design of Observational Studies*, New York: Springer. [1361,1368]

Rosenbaum, P. R., Ross, R. N., and Silber, J. H. (2007), "Minimum Distance Matched Sampling With Fine Balance in an Observational Study of Treatment for Ovarian Cancer," *Journal of the American Statistical Association*, 102, 75–83. [1362,1367]

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. [1366,1371]

——— (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 39, 33–38. [1361]

Rosenbaum, P. R., and Silber, J. (2001), "Matching and Thick Description in an Observational Study of Mortality After Surgery," *Biostatistics*, 2, 217–232. [1361]

Rubin, D. B. (1979), "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, 318–328. [1370]

Silber, J., Rosenbaum, P. R., Even-Shoshan, O., Mi, L., Kyle, F., Teng, Y., Bratzler, D., and Fleisher, L. A. (2011), "Estimating Anesthesia Time Using the Medicare Claim: A Validation Study," *Anesthesiology*, 115, 322–333. [1360]

Silber, J., Rosenbaum, P. R., Kelz, R. R., Reinke, C., Neuman, M., Ross, N., Even-Shoshan, O., David, G., Saynisch, P., Kyle, F., Bratzler, D., and Fleisher, L. A. (2012), "Medical and Financial Risks Associated With Surgery in the Elderly Obese," *Annals of Surgery*, 256, 79–86. [1360]

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods*, Ames, IA: Iowa State University Press. [1370]

Stevens, L. A., Coresh, J., Greene, T., and Levey, A. S. (2006), "Assessing Kidney Function–Measured and Estimated Glomerular Filtration Rate," *The New England Journal of Medicine*, 354, 2473–2483. [1360]

Stuart, E. A. (2010), "Matching Methods for Causal Inference: A Review and a Look Forward," *Statistical Science*, 25, 1–21. [1361,1366]

Yang, D., Small, D., 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. [1362,1367,1368]

Yoon, F. (2009), "Entire Matching and Its Application in an Observational Study of Treatments for Melanoma," Ph.D. dissertation, Department of Statistics, The Wharton School, University of Pennsylvania. [1370]

Zubizarreta, J. R., Reinke, C. E., Kelz, R. R., Silber, J. H., and Rosenbaum, P. R. (2011), "Matching for Several Sparse Nominal Variables in a Case-Control Study of Readmission Following Surgery," *The American Statistician*, 65, 229–238. [1362,1370]