

Comment: Matching Methods for Observational Studies Derived from Large Administrative Databases

Mark M. Fredrickson, Josh Errickson and Ben B. Hansen

1. INTRODUCTION

In the era of big data, finding a comparable control group for a set of treated units provides new opportunities and challenges. When controls vastly outnumber treated subjects, there will likely be many good potential matches for each treated subject. On the other hand, with larger data sets, increased computation time prevents applying existing methods to find the best possible match. Yu et al. propose a fast caliper solution that restricts the possible controls for each treated subject, making matching with large databases tractable. Their results on determining the narrowest caliper that is compatible with pair matching (without replacement) will be of great practical use.

We take issue with the labeling of this caliper as “optimal.” The label is accurate in a certain sense—it does minimize an objective of caliper width, subject to the constraint that pair matching remain feasible while no treatment group member is discarded—but these are quite different objectives and constraints from those otherwise targeted in the course of optimal matching. The meaning of “optimal” in “optimal matching” is already obscure to many, as Yu and Rosenbaum have themselves acknowledged (Yu and Rosenbaum, 2019). Adding a new and distinct connotation seems a step in the wrong direction.

It happens that Yu et al.’s optimal caliper can have the surprising result of forcing matches to be *suboptimal*, at least for the matching problem’s original objective. We demonstrate this phenomenon in a small stylized example (Section 2). Full matching is less affected; also, the narrowest caliper that is compatible with full matching is simple to describe and quick to calculate (Section 3). In another large surgical outcomes study, a form of full matching with restrictions is shown to generate matches

faster and with better optimality properties, while still maintaining a structure similar to pairs (Section 4). These critiques of pair matching notwithstanding, we expect caliper width to continue to be a leading determinant of matching speed, even as optimal matching in statistics assimilates algorithmic developments from related fields (Section 5).

2. OPTIMALITY AND THE EYE OF THE BEHOLDER

For a set of treated units \mathcal{T} and a set of control units \mathcal{C} , the match $\mathbf{m} = (m_{ij} : (i, j) \in \mathcal{T} \times \mathcal{C})^T$ has $m_{ij} = 1$ if treated unit i is matched to control unit j , 0 otherwise. The total distance of \mathbf{m} is given by

$$(1) \quad f(\mathbf{m}) = \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} m_{ij} d_{ij},$$

where d_{ij} is the distance between i and j . Yu et al. contribute to a body of work that uses careful application of network flow algorithms or advances in integer programming to minimize f (Rosenbaum, 1989, Hansen and Klopfer, 2006, Yang et al., 2012, Zubizarreta, 2012, Pimentel et al., 2015, Pimentel, Yoon and Keele, 2015, Rosenbaum, 2017, Pimentel et al., 2018). Yu et al. focus on pair matching, minimization of $f(\cdot)$ over

$$(2) \quad \left\{ \mathbf{m} \in \mathcal{T} \times \mathcal{C} : \sum_{j \in \mathcal{C}} m_{ij} = 1, \text{ all } i \in \mathcal{T}; \right. \\ \left. \sum_{i \in \mathcal{T}} m_{ij} \leq 1, \text{ all } j \in \mathcal{C} \right\}.$$

Matched sets corresponding to such \mathbf{m} maintain a strict 1 : 1 ratio of treated to control subjects. In contrast, *full matching* minimizes over strictly broader classes of \mathbf{m} , permitting both $i \in \mathcal{T}$ with $\sum_{j \in \mathcal{C}} m_{ij} > 1$ and $j \in \mathcal{C}$ with $\sum_{i \in \mathcal{T}} m_{ij} > 1$, while requiring that: for each $i \in \mathcal{T}$, $\sum_{j \in \mathcal{C}} m_{ij} \geq 1$; if $\sum_{j \in \mathcal{C}} m_{i'j} > 1$ then $\sum_{i \in \mathcal{T}} m_{ij'} = 1$ for each $j' \in \mathcal{C}$ s.t. $m_{i'j'} = 1$; and similarly if $\sum_{i \in \mathcal{T}} m_{ij'} > 1$ then $\sum_{j \in \mathcal{C}} m_{i'j} = 1$ for each $i' \in \mathcal{T}$ such that $m_{i'j'} = 1$ (Rosenbaum, 1991). That is, both many-one and one-many matched sets are permitted, as are 1 : 1 matched pairs; however, many-many configurations are excluded. Since optimal pair matching minimizes (1) over (2) and optimal full matching minimizes the same objective over a strictly broader domain, it is clear that for objective (1),

Mark M. Fredrickson is Postdoctoral Research Fellow, Department of Statistics, University of Michigan, 311 West Hall, 1085 South University Ave, Ann Arbor, MI 48109 (e-mail: mfredric@umich.edu). Josh Errickson is Statistician Senior, Consulting for Statistics, Computing and Analytics Research (CSCAR), University of Michigan 3550 Rackham Bldg, Ann Arbor MI 48109 (e-mail: jerrick@umich.edu). Ben B. Hansen is Associate Professor, Department of Statistics, University of Michigan, 323 West Hall, 1085 South University Ave, Ann Arbor, MI 48109 (e-mail: bbh@umich.edu).

full matching can be better than pair matching, but never worse.

In its unrestricted form, full matching suffers from other limitations. First, highly skewed matching ratios threaten efficiency (Hansen, 2004), as well as the simplicity that makes matched analysis uniquely accessible to nontechnical audiences (Rosenbaum and Rubin, 1985). This points toward full matching with *structural restrictions*, upper limits on the integers a for which $a : 1$ matches are permitted, and/or on integers b for which $1 : b$ matching ratios are allowed (Gu and Rosenbaum, 1993, Hansen and Klopfer, 2006). Second, as Yu et al. persuasively argue in the case of pair matching, computationally feasible matching in large databases calls for special measures (such as calipers) to reduce the number of $(i, j) \in \mathcal{T} \times \mathcal{C}$ for which the possibility that $m_{ij} \neq 0$ is entertained; we refer to this number as the *size* of the matching problem. Their argument applies with equal force to full matching, with or without structural restrictions.

Adding or subtracting any of these restrictions changes the nature of the matching problem, whether it be full or pair matching; generally one loses one form of optimality even as one gains another. Consider the pair matching problem, minimize (1) over (2), with $\mathcal{T} = \{A, B\}$, $\mathcal{C} = \{C, D, E\}$, and d_{ij} equal to the absolute difference of i and j on an underlying scalar measure for which $A = 0$, $B = 2$, $C = -3$, $D = 0$ and $E = 6$. Figure 1 visualizes these parameters in terms of an isomorphic minimum cost flow problem (Rosenbaum, 1989), set up for pair matching. In addition to the arcs between treated and control nodes, there are additional edges connecting these nodes to others at which flow enters or exits the system. Minimum cost flow analogues of full matching require additional nodes and structure (Rosenbaum, 1991), not shown here. Without further restrictions, the optimal pair match is $\{(A, D), (B, E)\}$ with a total distance of 4. With the addition of a caliper of 3, this match is no longer permitted. Pair matching remains feasible, but is forced to pair A to C in order to allow B its only permitted match to D ; total distance increases to 5. This is the smallest caliper compatible with pair matching, the one Yu et al. term “op-

timal”; yet it has precluded making the pair match that would otherwise be optimal.

3. FULL MATCHING: CALIPERS, RESTRICTIONS AND WEIGHTS

Adapting Yu et al.’s proposal from pair to full matching leads to an interestingly different result. A sufficient condition for full matching is that all $i \in \mathcal{T}$ have at least one $j \in \mathcal{C}$ for which $m_{ij} > 0$ is permitted. In the flow diagram analogous to that of Figure 1, each \mathcal{T} node must be connected to at least one \mathcal{C} node (Rosenbaum, 1991, Corollary 4). Thus the smallest caliper compatible with unrestricted full matching is the maximum over the treatment group of minimum distances to members of the control group.

PROPOSITION 1. *For an unrestricted full match, the minimum feasible caliper value c_t that discards no treated units is given by*

$$c_t = \max_{i \in \mathcal{T}} \min_{j \in \mathcal{C}} d_{ij},$$

where d_{ij} is the distance between treated unit i and control unit j .

If calculating a distance is a constant time operation, Proposition 1 gives a $O(|\mathcal{T}||\mathcal{C}|)$ time algorithm for finding the minimum feasible full match caliper. Proposition 1 does not place convexity or other conditions on the distances $\{d_{ij}\}$. In the situation where $d_{ij} = |v_i - v_j|$ for some scalar index v , the Algorithm 1 determines c_t in $O(|\mathcal{T}| \log(|\mathcal{T}|) + |\mathcal{C}| \log(|\mathcal{C}|))$ operations.

It may be the case that a caliper of c_t discards some control units. An analogous caliper, call it c_c , can be formed by switching the roles of treated and control in Proposition 1, and a sufficiently wide caliper to remove neither treated nor control group members would be $c = \max(c_t, c_c)$.

The caliper of Proposition 1 has the property that it serves as a lower bound on the minimum feasible caliper

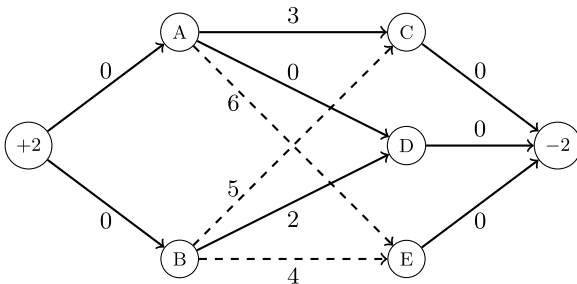


FIG. 1. Network flow diagram for a simple matching problem. All edge capacities are fixed at 1. Dashed lines would be eliminated for a caliper value of 3, the minimum feasible caliper for a pair match.

Algorithm 1 Minimum feasible full match caliper on scalar index v

```

Let  $u_i$  be the sorted  $v$  for the treated units.
Let  $w_j$  be the sorted  $v$  for the control units. Set  $w_0 = -\infty, w_{|\mathcal{C}|+1} = \infty$ .
 $c_t \leftarrow 0$ 
 $j \leftarrow 1$ 
for  $i = 1, \dots, |\mathcal{T}|$  do
  while  $w_j < u_i$  do
     $j \leftarrow j + 1$ 
  end while
   $c_t \leftarrow \max(c_t, \min(|u_i - w_{j-1}|, |u_i - w_j|))$ 
end for

```

for any restricted match, such as pair matching. A researcher seeking a narrow caliper, but committed to pair rather than full matching, can compute c_t and then check if the caliper is also feasible for pair matching. If not, the caliper size can be increased until a feasible caliper value is found for the pair match problem. Yu et al.'s adaptation of Glover's algorithm leads to the same destination, and is surely quicker, but this method is nicely flexible, as demonstrated in Section 4.

Working within narrow calipers, full matching may create highly imbalanced structures, sets with a large number a of treatment subjects sharing a single control or, alternately, sets pairing 1 treatment subject to $b \gg 1$ controls. The method of Yu et al. permits only 1 : 1 pairs, widening the caliper c_t just enough to make such matching feasible. As an alternative to pair matching, consider *full matching with symmetric restrictions*: matching in ratios of $k : 1$ up to $1 : k$, for some positive integer k , placing into matched sets each member of the treatment group and precisely as many controls as pair matching would have matched. Analogously to finding the Yu et al. optimal caliper for pair matching, the skewedness parameter k can be increased upward from 1 until just large enough for a caliper-respecting, symmetrically restricted full match to exist. By leaving the caliper fixed at c_t , this approach avoids both degrading the quality of matches, in the sense of (1), and increasing the number of potential matches to be stored in memory and then selected from among by the solver.

In post-matching analysis, it is relatively straightforward to account for pairs. Full matching, on the other hand, calls for analysis weights that vary with matched sets' sizes and treatment-control ratios. Under one frequently used weighting scheme, if a matched set has n_t treated units and n_c controls, they receive analysis weights $w_t = 2n_c/(n_t + n_c)$ and $w_c = 2n_t/(n_t + n_c)$, respectively. The factor of 2 ensures that when $n_t = n_c$, $w_t = w_c = 1$, recovering the natural weights for pair matching. This is the implicit weighting scheme of many common methods such as regression with fixed effects for matched sets, as well as tests based on sum statistics (Rosenbaum, 2002). Under various assumptions, it optimizes power (Kalton, 1968, Hansen and Bowers, 2008).

While a pair match will always have $|\mathcal{T}|$ pairs, the number of sets in a symmetrically restricted full match is not fixed *a priori*. One useful way to measure the size of a full match is the *effective sample size* (Hansen, 2011), which is the sum of w_t over all treated units (or equivalently the sum of w_c over all control units). This quantity expresses the information content of the matched sample (Kilcioglu and Zubizarreta, 2016) in terms of the number of similarly informative pairs that would be needed to equal it.

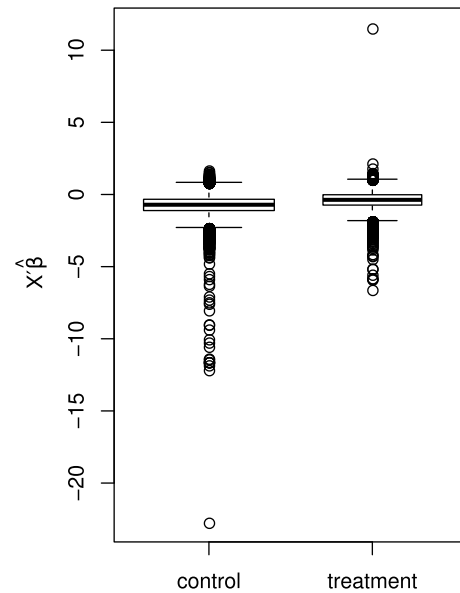


FIG. 2. Estimated propensity scores (logit scale) for patients undergoing percutaneous coronary intervention followed by manual closure of the arteriotomy site (“control”) versus closure assisted by a vascular closure device (“treatment”).

4. APPLICATION: OPTIMAL MATCHING TO OPTIMIZE SURGICAL PROCEDURE

Gurm et al. (2013) performed a propensity score matched analysis of the use of vascular closure devices (VCD) on patients undergoing percutaneous coronary intervention compared to patients for whom a VCD was not employed. As in Yu et al.'s surgical outcomes analysis of Medicaid data, the available sample was relatively large, with 31,000 treatment group members and 54,000 potential controls. Our reanalysis uses one of Gurm et al.'s three propensity scores, along with the same 192 exact matching categories respected by their match.

Figure 2 shows logits of estimated propensity scores. Two extreme outliers are present, one each in the treatment and control groups. While the control units with extreme values may raise little concern, particularly when the database of control units is large relative to the treated units, the extreme treated unit could lead to problems when picking a *single caliper for all units*. The nearest control in its exact matching group is separated from it by 10.3 logits, about 17 pooled standard deviations. A caliper of this size applied to all edges would remove only edges connected to the outlier, giving no meaningful reduction in matching problem size.

Design of observational studies typically involves restricting the analysis sample to a region of “common support,” explicitly or implicitly dropping outliers on the propensity score (Rosenbaum, 2002, Crump et al., 2009, Traskin and Small, 2011, Hill and Su, 2013, Imbens and Rubin, 2015, Fogarty et al., 2016), as a component or supplement of the matching process. Excluding the treatment outlier, Algorithm 1 finds the minimum feasible full

TABLE 1

Symmetrically restricted full match of the [Gurm et al. \(2013\)](#) sample. Matched set configurations and corresponding weights for treated (w_t) and control (w_c) units appear at left, with corresponding counts and effective sample size contributions at right (after rounding)

Design			Number of Sets	Effective Size
$n_t : n_c$	w_t	w_c		
3 : 1	0.5	1.5	2500	3800
2 : 1	0.7	1.3	2600	3500
1 : 1	1.0	1.0	12,900	12,900
1 : 2	1.3	0.7	2900	3900
1 : 3	1.5	0.5	2300	3500
Totals:			23,300	27,600

match caliper to be 0.48 logits. In addition to the treatment group outlier recognized in [Figure 2](#), another 51 members of the treatment group lacked control group counterparts within 0.48 logits that also shared their values of exact matching variables. We removed caliper restrictions on these, and only these, treatment group members, permitting them to match any control unit within their exact matching subclasses. (Our `optmatch` package offers a simple means of excluding designated units from a caliper, through an optional `exclude` argument that can be invoked in tandem with `caliper`.) The combination of caliper and exact matching categories decreased problem size by 99%.

Pair matching was not feasible with these caliper and exact matching requirements. Following the spirit of the Yu et al. caliper, we increased caliper width (as it applied to all but 52 members of the treatment group) in increments of 10%, stopping once pair matching became feasible. This process terminated after 18 iterates, with a caliper of width 1.344. This widening of the caliper increased matching problem size by a factor of 1.91.

Given the same caliper and exact matching requirements, optimal full matching produced 22,874 matched sets, 13,549 being 1 : 1 pairs but 690 having 5 or more units of one group matched to a single unit of the other. To avoid these imbalanced groupings, we explored feasibility of full matches with symmetric restrictions, beginning with $k = 1$ (pair matching) and then increasing k in increments of 1. This procedure terminated at $k = 3$. [Table 1](#) presents the structure of the resulting match.

One straightforward comparison of these different matches is in terms of $f(\cdot)$, the total of within set distances. The comparison, however, does not reflect the differing structures of the two matching strategies. To better reflect these differences, we normalize $f(\cdot)$ by the total number of treatment-control comparisons implied by the match (or equivalently, the number of arcs in the network diagram that carry flow from treated nodes to control nodes). Using this metric, the pair match had a total

(normalized) distance over 10 times larger than that of the restricted full match.

The pair match also required significantly more computation time. On a 2.30 GHz Intel Xeon E5-2630, producing the restricted full match required just under 16 minutes of computation time. Producing the pair match required nearly 30 minutes. This timing does not include the additional time required to find the parameters that lead to feasibility. Our approach required only two steps to find a feasible restricted full match. For the pair matching, the 10% steps required 18 increments to find a sufficiently wide caliper for pair matching.

5. OPTIMAL MATCHING'S FASTER FUTURE

While the paper of Yu et al. and this comment have focused on using calipers to make matching possible in large databases, we conclude with some notes on improving the matching algorithms themselves. Both the approach of Yu et al. and [Algorithm 1](#) used the special structure of the distance measure to find the minimum feasible calipers quickly. While network flow-based matching algorithms, such as the one used in `optmatch`, gain no additional benefits from this structure, other algorithms have similar time complexity in this setting ([Karp and Li, 1975](#), [Colannino et al., 2006](#), [Colannino et al., 2007](#), [Mohamad, Rappaport and Toussaint, 2015](#), [Rajabi-Alni and Bagheri, 2016](#)). If a caliper value—or a suitably smooth caliper function—is prespecified, [Ruzankin \(2019\)](#) provides fast algorithms that maximize the number of pairs, and also minimize (1) for many problems.

For general problems without such structure, we see great promise in parallelizing algorithms to make large matching problems tractable. While the relaxation-based algorithms that `optmatch` and `rcbalance` use do not lend themselves to parallelization, related “auction algorithms” do ([Bertsekas, 1998](#), Chapter 7), even admitting implementation on massively parallel GPUs ([Vasconcelos and Rosenhahn, 2009](#)). For sufficiently small treatment groups, multiple matches can be run in parallel and combined to produce globally optimal results ([Cohen and Brassil, 2000](#)). Slight modifications of the distance metric can also have computational benefits. [Sävje, Higgins and Sekhon \(2017\)](#) present a highly scalable variant of full matching optimizing the largest matched set diameter.

6. SUMMARY

Ruoqi Yu, Jeff Silber and Paul Rosenbaum have done the field an excellent service in demonstrating computational and other advantages of matching within narrow yet feasible calipers. Analysts should be aware that imposing any caliper, even an “optimal” caliper, can have the paradoxical effect of worsening match quality in the sense of

(1). Our case study underscored the need to deploy methods such as Yu et al.'s with attention to issues of common support. It also found full matching with symmetric restrictions to outperform pair matching significantly, in terms of both speed and quality of matching.

ACKNOWLEDGMENTS

We thank Ruoqi Yu, Jeffrey Silber and Paul Rosenbaum, for their engaging paper; the issue editors, Jason Roy and Dylan Small, for the opportunity to share our comments on it; and Joshua Buckner, Adam Rauh and Peter Solenberger, for contributions to the optmatch package enabling analysis methods we describe.

REFERENCES

- BERTSEKAS, D. (1998). *Network Optimization: Continuous and Discrete Models*. Athena Scientific, Belmont, MA.
- COHEN, N. and BRASSIL, J. (2000). A parallel pruning technique for highly asymmetric assignment problems. *IEEE Trans. Parallel Distrib. Syst.* **11** 550–558. <https://doi.org/10.1109/71.862206>
- COLANNINO, J., DAMIAN, M., HURTADO, F., IACONO, J., MEIJER, H., RAMASWAMI, S. and TOUSSAINT, G. (2006). An $O(n \log n)$ -time algorithm for the restriction scaffold assignment problem. *J. Comput. Biol.* **13** 979–989. [MR2231777 https://doi.org/10.1089/cmb.2006.13.979](https://doi.org/10.1089/cmb.2006.13.979)
- COLANNINO, J., DAMIAN, M., HURTADO, F., LANGERMAN, S., MEIJER, H., RAMASWAMI, S., SOUVAIN, D. and TOUSSAINT, G. (2007). Efficient many-to-many point matching in one dimension. *Graphs Combin.* **23** 169–178. [MR2320626 https://doi.org/10.1007/s00373-007-0714-3](https://doi.org/10.1007/s00373-007-0714-3)
- CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. and MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96** 187–199. [MR2482144 https://doi.org/10.1093/biomet/asn055](https://doi.org/10.1093/biomet/asn055)
- FOGARTY, C. B., MIKKELSEN, M. E., GAIESKI, D. F. and SMALL, D. S. (2016). Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *J. Amer. Statist. Assoc.* **111** 447–458. [MR3538678 https://doi.org/10.1080/01621459.2015.1112802](https://doi.org/10.1080/01621459.2015.1112802)
- GU, X. S. and ROSENBAUM, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *J. Comput. Graph. Statist.* **2** 405–420.
- GURM, H. S., HOSMAN, C., SHARE, D., MOSCUCCI, M. and HANSEN, B. B. (2013). Comparative safety of vascular closure devices and manual closure among patients having percutaneous coronary intervention. *Ann. Intern. Med.* **159** 660–666.
- HANSEN, B. B. (2004). Full matching in an observational study of coaching for the SAT. *J. Amer. Statist. Assoc.* **99** 609–618. [MR2086387 https://doi.org/10.1198/016214504000000647](https://doi.org/10.1198/016214504000000647)
- HANSEN, B. B. (2011). Propensity score matching to extract latent experiments from nonexperimental data: A case study. In *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland* (N. Dorans and S. Sinharay, eds.) 149–181. Springer.
- HANSEN, B. B. and BOWERS, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statist. Sci.* **23** 219–236. [MR2516821 https://doi.org/10.1214/08-STS254](https://doi.org/10.1214/08-STS254)
- HANSEN, B. B. and KLOPPER, S. O. (2006). Optimal full matching and related designs via network flows. *J. Comput. Graph. Statist.* **15** 609–627. [MR2280151 https://doi.org/10.1198/106186006X137047](https://doi.org/10.1198/106186006X137047)
- HILL, J. and SU, Y.-S. (2013). Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *Ann. Appl. Stat.* **7** 1386–1420. [MR3127952 https://doi.org/10.1214/13-AOAS630](https://doi.org/10.1214/13-AOAS630)
- IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference— for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. [MR3309951 https://doi.org/10.1017/CBO9781139025751](https://doi.org/10.1017/CBO9781139025751)
- KALTON, G. (1968). Standardization: A technique to control for extraneous variables. *Appl. Statist.* **17** 118–136. [MR0234599 https://doi.org/10.2307/2985676](https://doi.org/10.2307/2985676)
- KARP, R. M. and LI, S.-Y. R. (1975). Two special cases of the assignment problem. *Discrete Math.* **13** 129–142. [MR0384086 https://doi.org/10.1016/0012-365X\(75\)90014-X](https://doi.org/10.1016/0012-365X(75)90014-X)
- KILCIOGLU, C. and ZUBIZARRETA, J. R. (2016). Maximizing the information content of a balanced matched sample in a study of the economic performance of green buildings. *Ann. Appl. Stat.* **10** 1997–2020. [MR3592046 https://doi.org/10.1214/16-AOAS962](https://doi.org/10.1214/16-AOAS962)
- MOHAMAD, M., RAPPAPORT, D. and TOUSSAINT, G. (2015). Minimum many-to-many matchings for computing the distance between two sequences. *Graphs Combin.* **31** 1637–1648. [MR3386035 https://doi.org/10.1007/s00373-014-1467-4](https://doi.org/10.1007/s00373-014-1467-4)
- PIMENTEL, S. D., YOON, F. and KEELE, L. (2015). Variable-ratio matching with fine balance in a study of the Peer Health Exchange. *Stat. Med.* **34** 4070–4082. [MR3431322 https://doi.org/10.1002/sim.6593](https://doi.org/10.1002/sim.6593)
- 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. *J. Amer. Statist. Assoc.* **110** 515–527. [MR3367244 https://doi.org/10.1080/01621459.2014.997879](https://doi.org/10.1080/01621459.2014.997879)
- PIMENTEL, S. D., PAGE, L. C., LENARD, M. and KEELE, L. (2018). Optimal multilevel matching using network flows: An application to a summer reading intervention. *Ann. Appl. Stat.* **12** 1479–1505. [MR3852685 https://doi.org/10.1214/17-AOAS1118](https://doi.org/10.1214/17-AOAS1118)
- RAJABI-ALNI, F. and BAGHERI, A. (2016). An $O(n^2)$ algorithm for the limited-capacity many-to-many point matching in one dimension. *Algorithmica* **76** 381–400. [MR3537023 https://doi.org/10.1007/s00453-015-0044-4](https://doi.org/10.1007/s00453-015-0044-4)
- ROSENBAUM, P. R. (1989). Optimal matching for observational studies. *J. Amer. Statist. Assoc.* **84** 1024–1032.
- ROSENBAUM, P. R. (1991). A characterization of optimal designs for observational studies. *J. Roy. Statist. Soc. Ser. B* **53** 597–610. [MR1125717](https://doi.org/10.1111/14717)
- ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR1899138 https://doi.org/10.1007/978-1-4757-3692-2](https://doi.org/10.1007/978-1-4757-3692-2)
- ROSENBAUM, P. R. (2017). Imposing minimax and quantile constraints on optimal matching in observational studies. *J. Comput. Graph. Statist.* **26** 66–78. [MR3610408 https://doi.org/10.1080/10618600.2016.1152971](https://doi.org/10.1080/10618600.2016.1152971)
- ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33–38.
- RUZANKIN, P. S. (2019). A fast algorithm for maximal propensity score matching. *Methodol. Comput. Appl. Probab.* <https://doi.org/10.1007/s11009-019-09718-4>
- SÄVJE, F., HIGGINS, M. J. and SEKHON, J. S. (2017). Generalized full matching. arXiv e-prints, [arXiv:1703.03882](https://arxiv.org/abs/1703.03882).
- TRASKIN, M. and SMALL, D. S. (2011). Defining the study population for an observational study to ensure sufficient overlap: A tree approach. *Stat. Biosci.* **3** 94–118. <https://doi.org/10.1007/s12561-011-9036-3>

- VASCONCELOS, C. and ROSENHAHN, B. (2009). Bipartite graph matching computation on GPU. In *Energy Minimization Methods in Computer Vision and Pattern Recognition* (D. Cremers, Y. Boykov, A. Blake and F. R. Schmidt, eds.). *Lecture Notes in Computer Science* **5681** 42–55. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03641-5_4
- 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. MR2959630 <https://doi.org/10.1111/j.1541-0420.2011.01691.x>
- YU, R. and ROSENBAUM, P. R. (2019). Directional penalties for optimal matching in observational studies. *Biometrics* **75** 1380–1390. MR4041838 <https://doi.org/10.1111/biom.13098>
- ZUBIZARRETA, J. R. (2012). Using mixed integer programming for matching in an observational study of kidney failure after surgery. *J. Amer. Statist. Assoc.* **107** 1360–1371. MR3036400 <https://doi.org/10.1080/01621459.2012.703874>