



Published in final edited form as:

*J Comput Graph Stat.* 2008 December ; 17(4): 914–924. doi:10.1198/106186008X385806.

## An Algorithm for Optimal Tapered Matching, With Application to Disparities in Survival

Shoshana R. Daniel, Katrina Armstrong, Jeffrey H. Silber, and Paul R. Rosenbaum<sup>1</sup>

University of Pennsylvania, Philadelphia

### Abstract

In a tapered matched comparison, one group of individuals, called the focal group, is compared to two or more nonoverlapping matched comparison groups constructed from one population in such a way that successive comparison groups increasingly resemble the focal group. An optimally tapered matching solves two problems simultaneously: it optimally divides the single comparison population into nonoverlapping comparison groups and optimally pairs members of the focal group with members of each comparison group. We show how to use the optimal assignment algorithm in a new way to solve the optimally tapered matching problem, with implementation in R. This issue often arises in studies of groups Defined by race, gender, or other categorizations such that equitable public policy might require an understanding of the mechanisms that produce disparate outcomes, where certain specific mechanisms would be judged illegitimate, necessitating reform. In particular, we use data from Medicare and the SEER Program of the National Cancer Institute as part of an ongoing study of black-white disparities in survival among women with endometrial cancer.

### Keywords

Assignment algorithm; combinatorial optimization; matched sampling; tapered matching

## 1 Introduction and Review

### 1.1 Introduction: What is Tapered Matching?

In a tapered matched comparison, one group of individuals, called the focal group, is compared to two or more nonoverlapping matched comparison groups constructed from one population in such a way that successive comparison groups increasingly resemble the focal group. ‘Tapering’ refers to this narrowing or successive removal of naturally occurring discrepancies between the focal group and the comparison population. The focal group never changes; it represents the focal group as it actually exists in the population. The successive comparison groups resemble their own population less and less, and resemble the focal population more and more. Tapered matching is used in an effort to understand how and why outcomes in the focal population differ from those in the comparison population, by successively removing some discrepancies that may be responsible for part of the

<sup>1</sup>Address for correspondence: Department of Statistics, The Wharton School, University of Pennsylvania, 473 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340 USA. rosenbaum@stat.wharton.upenn.edu.

difference in outcomes. Because the comparison groups are nonoverlapping, they may be compared to one another using simple statistical methods. Tapered comparisons of this sort shed some light on questions of the form: When certain baseline discrepancies are removed, how much of the disparity in outcomes is also removed?

An optimally tapered matching solves two problems simultaneously: it optimally divides the single population into  $C - 2$  nonoverlapping comparison groups and optimally pairs members of the focal group with members of each comparison group. More generally, each member of the focal group may be matched to  $k_c - 1$  controls in comparison group  $c$ ,  $c = 1, \dots, C$ . For discussion of matching with one comparison group and multiple controls from that group, see Smith (1997).

This problem occurs in various contexts, including studies of disparities in outcomes in which one is seeking insight into mechanisms that produce the disparities. In the example, black women with endometrial cancer (the focal group) survive for a shorter period following diagnosis than do white women, and we compare black women to white women who were diagnosed with similar cancer (a wider comparison group) and white women who were similar in diagnosis and also similar in surgical treatment (a narrower comparison group). Because the two white groups do not overlap, they may be compared to each other using conventional statistical methods to ask: If whites who were diagnosed in the same health as blacks also received the same surgery as blacks, could the difference in surgery help in understanding the difference in survival? Unaided by matching, a model fitted to the entire population would give disproportionate weight to the large and comparatively healthy white population, and that model may substantially misrepresent the situation in the small, sicker black population; see Dehejia and Wahba (1999) for discussion of the failure of models unaided by matching. In contrast, it is often useful to apply model-based adjustments to matched samples (Rubin 1979).

We use the assignment algorithm in a new way to solve the optimally tapered matching problem. The assignment algorithm is reviewed in §1.2. An example is discussed in §2. Advantages of tapered matching in the context of the example are discussed in §2.2. The method for optimal tapered matching is discussed in §3. Technical details of the example in §2 are discussed in §3.3. The choice of distance function in tapered matching is discussed in §3.4.

## 1.2 Review: The Optimal Assignment Algorithm

The assignment algorithm is one of the oldest combinatorial optimization algorithms, and it solves the following problem. There is an  $A \times B$  matrix of nonnegative, possibly infinite ‘distances,’  $\delta_{ab} \geq 0$ ,  $a = 1, \dots, A$ ,  $b = 1, \dots, B$ , with  $A \leq B$ . These ‘distances’ need not satisfy the triangle inequality, so they need not be distances in the sense of a metric space. One familiar distance is the Mahalanobis distance (Rubin 1979). An *assignment*,  $\beta(\cdot)$ , of columns to rows pairs each row  $a \in \{1, \dots, A\}$  to a different column  $\beta(a) \in \{1, \dots, B\}$  with  $\beta(a) \neq \beta(a')$  if  $a \neq a'$ . Let  $\mathcal{B}$  be the set containing all  $B!/ (B-A)!$  assignments  $\beta(\cdot)$ . The *total distance*,  $\delta\{\beta(\cdot)\}$ , for an assignment  $\beta(\cdot)$  is sum of the distances for its  $A$  row-column pairs,

$\delta\{\beta(\cdot)\} = \sum_{a=1}^A \delta_{a,\beta(a)}$ . The *optimal assignment problem* is to find any assignment,  $\tilde{\beta}(\cdot) \in \mathcal{B}$ , that is optimal in the sense of minimizing the total distance over all assignments:

$$\delta\{\tilde{\beta}(\cdot)\} = \min_{\beta(\cdot) \in \mathcal{B}} \delta\{\beta(\cdot)\} \text{ or equivalently } \sum_{a=1}^A \delta_{a,\tilde{\beta}(a)} = \min_{\beta(\cdot) \in \mathcal{B}} \sum_{a=1}^A \delta_{a,\beta(a)} \quad (1)$$

The problem is not trivial: two or more rows,  $a$  and  $a'$ , may want the same column  $b$  — that is,  $\delta_{ab} = \min_j \delta_{aj}$  and  $\delta_{a'b} = \min_j \delta_{a'j}$  — so the  $A$  assignments of columns to rows are interdependent. The assignment algorithm and related procedures have been used for optimal matching in observational studies; see Rosenbaum (1989, 1991), Silber, et al. (2001, 2005, 2007), Hansen (2004, 2007), Hansen and Klopfer (2006), Lu (2005), Augurzky and Kluge (2007) and Rosenbaum, Silber and Ross (2007). One of the fastest algorithms holds an auction, in which two rows  $a$  and  $a'$  that both want the same column  $b$  offer competing bids for it, with the bids adjusting as the auction progresses; see Bertsekas (1981). The pairmatch function in Hansen's (2004, 2007) optmatch package in R makes Bertsekas' very fast Fortran code available from within R. Dell'Amico and Toth (2000) review and compare algorithms to solve the assignment problem. As discussed by Papadimitriou and Steiglitz (1982), polynomial time algorithms for the assignment problem are available, including Kuhn's Hungarian algorithm, with time bound of  $O(B^3)$ . Carpaneto, and Toth (1980) provide Fortran code implementing the Hungarian method. In SAS OR, Proc Assign solves the assignment problem; see Bergstralh, Kosanke and Jacobsen (1996).

As we demonstrate in §3, for a suitably Defined matrix , the problem of optimal tapered matching can be reduced to the assignment problem. Specifically, the assignment algorithm will both partition the comparison population into nonoverlapping comparison groups matched in varying degrees to the focal group and also match each member of the focal group to  $k_c$  similar members of comparison group  $c$ ,  $c = 1, \dots, C$ . To make the discussion tangible, before discussing the procedure, its application is illustrated in §2 using the example that motivated our work.

## 2 Example: Disparities in Survival with Endometrial Cancer

### 2.1 Matching At Diagnosis and At Treatment: An Application of Tapered Matching

In the US Medicare population, black women with endometrial cancer survive for a shorter time following diagnosis than do white women. Why? Biological differences are not inconceivable, but there is naturally concern that the health care system might be providing inferior care to black women. In principle, all women in the Medicare population have equal access to medical care, but gaps between principle and practice may occur. We examine this using merged data from Medicare and from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, which provides clinical stage, grade and histology.

We begin with the population of black women diagnosed with endometrial cancer, specifically 806 black women diagnosed between 1991 and 2000. We compare them to two optimally matched, non-overlapping comparison groups: (i) white women who appeared to

have similar cancer at diagnosis, and (ii) white women who appeared to have similar cancer at diagnosis and who received similar surgical treatment. The construction of the matched comparison is discussed in general in §3 with specifics for this example in §3.3. We also compare with the unmatched population of white women with endometrial cancer. This offers several perspectives and answers several questions. Are blacks and whites diagnosed with similarly advanced cancer? (They aren't.) Do blacks and whites with similar cancer at diagnosis have similar survival? (They don't.) Do blacks and whites with similar cancer receive the same surgery? (They don't.) Do blacks and whites with similar cancer who receive the same surgery exhibit similar survival? (They don't.) Do whites who receive the same inferior treatment as blacks have poorer survival than other whites with similar cancer at diagnosis? (No, or at least there is no indication that they do.)<sup>2</sup>

Table 1 describes the four groups of women diagnosed with endometrial cancer between 1991 and 2000: (i) all 806 black women, (ii) 806 matched white women who appeared to have similar cancer at diagnosis, (iii) 806 other matched white women who appeared to have similar cancer at diagnosis and received similar surgical treatment, and (iv) all 13,756 white women. Both matched groups were matched for age, clinical stage or missing clinical stage, grade or missing grade, histology or missing histology, and the presence or absence of comorbid conditions; these variables are described in the top half of Table 1. In addition, the whites matched at treatment were matched for adequate surgery (AS = yes or no), surgeon type (GO = gynecological oncologist, GYN = gynecologist, Other = other surgeon), and an approximate measure of days from diagnosis to surgery, where surgeon type and days are defined only for the subset of women who received adequate surgery; these variables are described in the bottom half of Table 1.

In Table 1, black women were sicker at diagnosis than all white women. Black women tended to have more advanced stage of cancer, higher grade, and were more likely to present with a comorbid condition. In contrast, both matched groups of white women were similar to the black women in terms of the variables in the top half of Table 1. Black women received inferior medical care when compared to all white women and to the first matched group of white women who appeared similar in terms of the biological variables in Table 1. Black women were more likely to receive either no surgery or inadequate surgery (35% for blacks, 18% for similar whites matched at diagnosis, and 14% for the healthier group of all whites.) In contrast, the second matched group of white women was similar to the black women in terms of both the biological variables in Table 1 and the treatment variables in Table 2.

Comparisons of survival in matched pairs are based on Albers (1988) rank test for randomly censored matched pairs, using his Wilcoxon-type scores. Largely because blacks were diagnosed with more advanced cancer, blacks typically survived for a much shorter time than the population of all whites (median 2.9 years for blacks, with 95% confidence interval [2.4, 3.3] versus 8.7 years for all whites, [8.4, 8.9]). White women matched for health at

---

<sup>2</sup>The analyses presented here were a pilot study demonstrating the feasibility of the methodology. A clinical analysis along the same lines is currently underway; it will differ in a number of particulars, including greater attention to the extent and nature of comorbid conditions. A reader interested in endometrial cancer rather than in multivariate matching should refer to our clinical analysis.

diagnosis survived somewhat longer than the black women (median 4.1 years, [3.5, 4.8], with Albers' one-sided p-value 0.037). Recall from Table 1 that this first matched group of white women was more likely to receive adequate surgery than the group of black women. When matched also for surgery, the second matched group of white women still had somewhat longer survival than the black women (median 3.6 years, [2.7, 4.3], with Albers' one-sided p-value 0.033). Survival in the two nonoverlapping matched groups of white women did not differ significantly, despite the higher frequency of adequate surgery in the first matched group (Albers' one-sided p-value 0.48).

In short, blacks are diagnosed with more advanced cancer, and this might account for much of the difference in survival. Blacks are less likely to receive adequate surgery than matched whites who were similar at diagnosis, but there is little indication that this difference in surgery is the cause of the remaining difference in survival, because whites who received similar surgical treatment still survived longer. These observations reflect the covariates in Table 1, and might be different with more detailed clinical data.

## 2.2 Advantages of Tapered Matching in this Example

The example in §2.1 illustrates several advantages of tapered matching.

**Focus on the correct population**—The black population is the natural one: (essentially) all blacks diagnosed with endometrial cancer in the SEER-Medicare cohort, and that population does not change as the analysis proceeds. All blacks are compared to whites who become more and more similar to the blacks, but the comparisons always refer to, and are weighted by, the entire actual population of blacks. The question, “How beneficial is adequate surgery?,” is a different question if answered in the black population than in the entire population, which is mostly white and mostly diagnosed at an earlier, healthier stage. The analysis in §2.1 suggests substantial improvements in survival for blacks are mostly likely to be achieved through earlier diagnosis, despite the fact that there is also a disparity in the use of adequate surgery. See Dehejia and Wahba (1999) for discussion of the dangers of model-based adjustments, unaided by matching, which may not focus on the correct population.

**Adjusting for interactions**—In clinical data, the very meaning of one matching variable often changes depending upon the value of another matching variable, and for this reason it is important to control for interactions. The stage of endometrial cancer can be determined approximately by biopsy, more accurately as a by-product of adequate surgery, and still more accurately as a by-product of accurate surgery by a gynecological oncologist whose specialty is gynecological cancer surgery. Because of this, the meaning of “stage 3” is different for a woman who had adequate surgery than for one who did not, and possibly different depending upon the surgeon type (Earle, et al. 2006, Silber, et al. 2007). Time to surgery and surgeon type in Table 1 have meaning only for a patient who has surgery. A “missing stage” is unlikely to be missing completely at random, as it would result if even a biopsy was not performed; see the appendix of Rosenbaum and Rubin (1984) for the relationship between matching and incomplete covariates. In Table 1, blacks lacking key clinical data are matched to whites lacking the same data, and whites matched at treatment

are matched to blacks for ‘adequate surgery’ and typically for surgeon type, so in these black-white pairs, important interactions are controlled.

**The comparison is easy to understand**—Although matching itself requires some technology, in the end, patients who are comparable in specific respects are compared in a straightforward way. Anyone can understand the degree and nature of the comparability displayed in Table 1, and the comparison of survival outcomes. If scientific findings are to affect policy, it is helpful to have an analysis that nonstatisticians can understand.

### 3 Optimal Tapered Matching Using the Assignment Algorithm

#### 3.1 The Tapered Matching Problem

The focal group consists of  $I$  individuals,  $\alpha_1, \dots, \alpha_I$ , and the comparison population consists of  $J$  individuals,  $\gamma_1, \dots, \gamma_J$ . In §2,  $I = 806$  and  $J = 13,756$ . A total of  $C = 2$  tapered, nonoverlapping comparison groups will be formed,  $c = 1, \dots, C$ , and  $\alpha_i$  will be matched to  $k_c - 1$  individuals in comparison group  $c$  such that each of the  $J$  individuals  $\gamma_1, \dots, \gamma_J$  appears at most once among the  $T = IK$  matched controls, where  $K = \sum k_c$ . In §2,  $C = 2$ ,  $k_1 = k_2 = 1$ ,  $K = 2$ ,  $T = IK = 1612$ . Matching is possible only if  $J \geq T$ . Let  $|H|$  denote the number of elements of a finite set  $H$ .

The variables used in forming comparison group  $c$  are different from the variables used in forming comparison group  $c'$  for  $c' \neq c$ , so the distance between  $\alpha_i$  and  $\gamma_j$  changes with  $c$ . In §2, the variables used in forming comparison group  $c = 1$  described the health of patients at diagnosis, whereas the variables used in forming group  $c = 2$  described both the health of patients at diagnosis and their surgical treatment. Let  $\lambda_{cij} \geq 0$  be the distance between  $\alpha_i$  and  $\gamma_j$  when forming group  $c$ . In §2,  $\lambda_{1ij}$  is the distance between black woman  $\alpha_i$  and white woman  $\gamma_j$  in terms of variables describing health at diagnosis, whereas  $\lambda_{2ij}$  describes the distance between these same two women in terms of health at diagnosis and surgical treatment. Two women,  $\alpha_i$  and  $\gamma_j$ , might have been similar in terms of health at diagnosis, so  $\lambda_{1ij}$  is small, but they might have been treated very differently, so  $\lambda_{2ij}$  is large. An infinite distance,  $\lambda_{cij} = \infty$ , is used to prevent matching  $\alpha_i$  and  $\gamma_j$  at level  $c$  of the taper. In §2, we insisted on an exact match on clinical stage at both levels of the taper,  $c = 1, 2$ , and an exact match on the binary indicator of adequate surgery at level  $c = 2$  of the taper, so  $\lambda_{cij} = \infty$  if  $\alpha_i$  and  $\gamma_j$  differed in these ways.

The problem is to determine the subset of  $\{\gamma_1, \dots, \gamma_J\}$  to be matched to  $\alpha_i$  at each level  $c$  of the taper. Let  $\omega_{ci} \subset \{\gamma_1, \dots, \gamma_J\}$  be the set containing the  $k_c$  individuals matched to  $\alpha_i$  at level  $c$  of the taper; so  $\omega_{ci}$  identifies  $k_c$  controls,  $|\omega_{ci}| = k_c$ , and no control is used twice,  $\omega_{ci} \cap \omega_{c'i'} = \emptyset$  if  $i' \neq i$  or  $c' \neq c$ , and there are  $T = IK$  distinct controls in all,

$T = IK = \left| \bigcup_{i=1}^I \bigcup_{c=1}^C \omega_{ci} \right|$ . Write  $\Omega$  for the  $I \times C$  array of sets  $\omega_{ci}$ ,  $i = 1, \dots, I$ ,  $c = 1, \dots, C$ . The optimal tapered matching problem, Problem 1, is to construct  $\Omega$  in such a way as to minimize the total of the distances  $\lambda_{cij}$  between focal group members,  $\alpha_i$ , and their matched controls,  $\gamma_j, j \in \omega_{ci}$ .

**Problem 1** An  $I \times C$  array  $\Omega$  of sets  $\omega_{ci} \subset \{\gamma_1, \dots, \gamma_J\}$ ,  $i = 1, \dots, I$ ,  $c = 1, \dots, C$ , such that  $|\omega_{ci}| = k_c$ ,  $\omega_{ci} \cap \omega_{c'i'} = \emptyset$  if  $i' \neq i$  or  $c' \neq c$  is called feasible. Find a feasible array  $\Omega$  that minimizes  $\lambda(\Omega) = \sum_{i=1}^I \sum_{c=1}^C \sum_{j \in \omega_{ci}} \lambda_{cij}$ .

### 3.2 An Equivalent Assignment Problem

Define a matrix with  $T = IK$  rows and  $J$  columns as follows. The  $J$  columns correspond with  $\gamma_1, \dots, \gamma_J$ . The  $IK$  rows divide into  $I$  groups of  $K$  rows, one group for each  $\alpha_i$ . The  $K = \sum k_c$  rows in each of the  $I$  groups further divide into  $k_1$  rows for the first level of the taper,  $c = 1$ ,  $k_2$  rows for the second level,  $c = 2, \dots, k_C$  rows for level  $C$ . Form by placing  $\lambda_{cij}$  in the  $k_c$  rows

$$(i-1)K + \sum_{b=1}^{c-1} k_b + 1, \dots, (i-1)K + \sum_{b=1}^{c-1} k_b + k_c, \quad (2)$$

and column  $j$ , for  $i = 1, \dots, I$ ,  $c = 1, \dots, C$ ,  $j = 1, \dots, J$ , with  $\sum_{b=1}^0 k_b = 0$ .

**Method 1:** Solve the optimal assignment problem for the distance matrix. For the resulting optimal assignment, Define  $\omega_{ci}$  to be the columns paired with the rows for  $\alpha_i$  with the  $k_c$  indices (2), for  $i = 1, \dots, I$ ,  $c = 1, \dots, C$ .

**Proposition 2** The tapered matching method solves Problem 1

**Proof.** Each  $\beta(\cdot) \in \mathcal{B}$  determines a feasible  $\Omega$  in Problem 1 using Method 1; then, the  $\omega_{ci}$  so Defined are disjoint because  $\beta(\cdot)$  is an assignment and have  $|\omega_{ci}| = k_c$  by construction. Conversely, if  $\Omega$  is any feasible array of  $\omega_{ci}$  in Problem 1, then  $\Omega$  picks out  $T = IK$  distinct columns of from  $\{\gamma_1, \dots, \gamma_J\}$  because the  $\omega_{ci}$  do not overlap. For each  $c$  and  $i$ , there are  $k_c!$  ways to assign the columns in  $\omega_{ci}$  to the rows of with indices (2), so each  $\Omega$  corresponds with  $(\prod_{c=1}^C k_c!)$  different assignments  $\beta(\cdot) \in \mathcal{B}$  with the same cost  $\delta\{\beta(\cdot)\}$  which equals  $\lambda(\Omega)$ . It follows that any minimum cost assignment  $\beta(\cdot)$  corresponds with a minimum cost feasible array  $\Omega$ .

In Hansen's (2007) optmatch package, when  $k_1 = \dots = k_C > 1$ , a smaller distance matrix with  $IC$  rows may be used in place of in (2) with  $IK - IC$  rows, by instructing the package to match each row to  $k_1 > 1$  controls. In R, an unnecessarily large distance matrix may lead to an inefficient use of memory, with slower performance.

### 3.3 Details of Implementation in the Example

The optimization used Hansen's (2005, 2007) optmatch package in R. The two distances,  $\lambda_{cij}$ ,  $c = 1, 2$ , were each formed using Mahalanobis metric matching within calipers on two propensity scores. The two distances,  $\lambda_{cij}$ ,  $c = 1, 2$ , for matching at the time of diagnosis,  $c = 1$ , and at the time of treatment,  $c = 2$ , were Defined as follows. For the diagnosis match, the probability that a patient is black rather than white was estimated from the variables in the top half of Table 1 using a logit model; for matching purposes, this is similar to a propensity score. For the treatment match, the probability that a patient is black rather than white was

estimated from all of the variables in Tables 1 using a logit model. For discussion of the consequences of using an estimated propensity score, see Hirano, et al. (2003). In both matches, we insisted on an exact match on clinical stage or stage missing. This can be implemented with infinite distances, but in fact it is more efficient in time and space to divide the matching problem into five separate problems, one for each level of stage. The distance  $\lambda_{1ij}$  was calculated as the Mahalanobis distances (Rubin 1979) for the variables in the top of Table 1, and  $\lambda_{2ij}$  was calculated as the Mahalanobis distances for all the variables in Tables 1, with certain penalties added. Specifically, a penalty of  $10^6$  was added to  $\lambda_{2ij}$  if  $\alpha_i$  and  $\gamma_j$  differed on the binary indicator of adequate surgery; this ensured that in the treatment match,  $c = 2$ , patients were exactly matched for adequate surgery. In addition, penalties of  $10^3$  were added to  $\lambda_{cij}$  if  $\alpha_i$  and  $\gamma_j$  differed on propensity score  $c$  by more than one fifth of a standard deviation of the propensity score  $c$ ,  $c = 1, 2$ , and a penalty of  $10^3$  was added if  $\alpha_i$  and  $\gamma_j$  differed on the binary indicator for comorbidity.

This use of penalties is standard in optimization: it replaces a constraint by a change in the objective function. With a sufficiently large penalty, if the constraint can be respected, it will be, and if the constraint cannot be respected in every case, it will be respected as often as possible. In Table 1, in the match at diagnosis, the comorbidity constraint was respected in  $796/806 = 98.8\%$  of pairs. In the match at treatment, there are  $805/806 = 99.9\%$  pairs matched for comorbidity.

### 3.4 Choice of Distance

In tapered matching, two or more definitions of ‘distance,’ say  $\lambda_{1ij}$  and  $\lambda_{2ij}$ , are added together in  $\lambda(\Omega)$ . These different distances need to be such that they can reasonably be added together. Use of the Mahalanobis distance, with or without added penalties, is attractive in this context. This brief section explains why.

Suppose that the covariates for the  $I + J$  individuals  $\{\alpha_1, \dots, \alpha_I, \gamma_1, \dots, \gamma_J\}$  are in a matrix  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_C)$  with  $I + J$  rows and  $L = \ell_1 + \dots + \ell_C$  columns, where level  $c = 1$  of the taper will match on the  $\ell_1$  variables in  $\mathbf{V}_1 = \mathbf{X}_1$ , level  $c = 2$  of the taper will match on the  $\ell_1 + \ell_2$  variables in  $\mathbf{V}_2 = (\mathbf{X}_1, \mathbf{X}_2)$ , and generally, level  $c$  of the taper will match on the  $\ell_1 + \dots + \ell_c$  variables in  $\mathbf{V}_c = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c)$ , for  $c = 1, \dots, C$ . Here, row  $i$  of  $\mathbf{V}_c$ , say  $\mathbf{v}_{ci}$ , describes  $\alpha_i$ ,  $i = 1, \dots, I$ , while row  $I + j$  of  $\mathbf{V}_c$ , say  $\mathbf{v}_{c,I+j}$ , describes  $\gamma_j$ ,  $j = 1, \dots, J$  and each is of dimension  $\ell_1 + \dots + \ell_c$ . Suppose that  $\lambda_{cij}$  is the sample Mahalanobis distance between  $\alpha_i$  and

$\gamma_j$  computed from  $\mathbf{V}_c = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c)$ , that is,  $\lambda_{cij} = (\mathbf{v}_{ci} - \mathbf{v}_{c,I+j}) \hat{\Sigma}_c^{-1} (\mathbf{v}_{ci} - \mathbf{v}_{c,I+j})^T$ ,

where  $\hat{\Sigma}_c$  is the sample covariance matrix computed from  $\mathbf{V}_c$ . The Mahalanobis distances are invariant under affine transformations of the data matrix. In particular, if Gram-Schmidt orthogonalization (Rao 1973, §1b.2ix) were applied to the columns of  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_C)$ , and the resulting columns were standardized to mean zero and unit variance, and the Mahalanobis distances, the  $\lambda_{cij}$ 's, were recomputed from these uncorrelated, standardized variables, then all  $C \times I \times J$  distances  $\lambda_{cij}$  would be unchanged. In other words, without loss of generality, one may view all  $C \times I \times J$  Mahalanobis distances  $\lambda_{cij}$  as having been computed from uncorrelated variables with sample mean zero and sample standard deviation one. For such variables, the Mahalanobis distance  $\lambda_{cij}$  is the sum of  $\ell_1 + \dots + \ell_c$  squared



differences in the variables describing  $\alpha_i$  and  $\gamma_j$ . In other words, in the criterion  $\lambda(\Omega)$ , a difference of one standard deviation in one of these standardized variables contributes a value of one no matter where it occurs. The first variable in Table 1 is age. If the standard deviation of age in this Medicare population were five years, and if  $\alpha_i$  were 70 and  $\gamma_j$  were 75 then  $\alpha_i$  and  $\gamma_j$  would differ by one standard deviation. Speaking informally, the two Mahalanobis distances count this five year difference in age between  $\alpha_i$  and  $\gamma_j$  as equally serious for the match at diagnosis and the match at treatment, even though there are more variables in the match at treatment, so the distances tend to be larger in total at treatment. Recall that Gram-Schmidt orthogonalization makes the columns of  $\mathbf{X}$  orthogonal by, essentially, finding the residuals when column  $k$  of  $\mathbf{X}$  is regressed on columns 1,  $\dots$ ,  $k-1$ , for  $k=2, \dots, L$ . Speaking somewhat more formally, the Mahalanobis distance between  $\alpha_i$  and  $\gamma_j$  at treatment, namely  $\lambda_{2ij}$ , equals the Mahalanobis distance between  $\alpha_i$  and  $\gamma_j$  at diagnosis, namely  $\lambda_{1ij}$ , plus the Mahalanobis distance between  $\alpha_i$  and  $\gamma_j$  on the residuals of the treatment variables after regression on the diagnostic variables. In consequence, the difference between  $\alpha_i$  and  $\gamma_j$  on the diagnostic variables in Table 1 counts the same, namely  $\lambda_{1ij}$ , at both levels of the taper, but there is an additional contribution from the treatment variables, so  $\lambda_{2ij} > \lambda_{1ij}$ . In words, the Mahalanobis distance does not regard mismatches of a certain magnitude on the diagnostic variables to be more or less important when matching also on the surgical variables, and this is appropriate if both groups are to be well-matched in Table 1. Similar considerations apply to the large penalties, whether they are added because of violations of the propensity score caliper or because of mismatches on a specific variable such as comorbidity.

## Acknowledgments

This work was supported by grants from the Measurement, Methodology and Statistics Program of the U.S. National Science Foundation and from the U.S. National Cancer Institute.

## References

1. Albers W. Combined rank tests for randomly censored paired data. *Journal of the American Statistical Association*. 1988; 83:1159–1161.
2. Augurzky B, Kluve J. Assessing the performance of matching algorithms when selection into treatment is strong. *Journal of Applied Econometrics*. 2007; 22:533–557.
3. Bergstralh EJ, Kosanke JL, Jacobsen SL. Software for optimal matching in observational studies. *Epidemiology*. 1996; 7:331–332. SAS code is available at: <http://www.mayo.edu/hsr/sasmac.html>. [PubMed: 8728456]
4. Bertsekas DP. A new algorithm for the assignment problem. *Mathematical Programming*. 1981; 21:152–171. Fortran code is available at: <http://web.mit.edu/dimitrib/www/home.html>.
5. Carpaneto G, Toth P. Algorithm 548: Solution of the Assignment Problem [H]. *ACM Transactions on Mathematical Software*. 1980; 6:104–111. Fortran code is available at <http://portal.acm.org>.
6. Dehejia R, Wahba S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association*. 1999; 94:1053–1062.
7. Dell'Amico M, Toth P. Algorithms and codes for dense assignment problems: the state of the art. *Discrete Applied Mathematics*. 2000; 100:17–48.
8. Earle CC, Schrag D, Neville BA, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. *Journal of the National Cancer Institute*. 2006; 98:172–180. [PubMed: 16449677]

9. Hansen BB. Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*. 2004; 99:609–618.
10. Hansen BB, Klopfer SO. Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*. 2006; 15:609–627.
11. Hansen BB. Optmatch: flexible, optimal matching for observational studies. *R News*. 2007; 7:18–24.
12. Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. 2003; 71:1161–1189.
13. Lu B. Propensity score matching with time-dependent covariates. *Biometrics*. 2005; 61:721–728. [PubMed: 16135023]
14. Papadimitriou, C.; Steiglitz, K. *Combinatorial Optimization*. Englewood Cliffs, NJ: Prentice-Hall; 1982.
15. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation; 2007. <http://www.R-project.org>
16. Rao, CR. *Linear Statistical Inference and Applications*. NY: Wiley; 1973.
17. Rosenbaum PR. Optimal matching in observational studies. *Journal of the American Statistical Association*. 1989; 84:1024–1032.
18. Rosenbaum PR. A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society, B*. 1991; 53:597–610.
19. Rosenbaum P, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984; 79:516–524.
20. Rosenbaum PR, Ross RN, Silber JH. Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *Journal of the American Statistical Association*. 2007; 102:75–83.
21. Rubin DB. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*. 1979; 74:318–328.
22. Silber JH, Rosenbaum PR, Trudeau ME, Even-Shoshan O, Chen W, Zhang X, Mosher RE. Multivariate matching and bias reduction in the surgical outcomes study. *Medical Care*. 2001; 39:1048–1064. [PubMed: 11567168]
23. Silber JH, Rosenbaum PR, Trudeau ME, Chen W, Zhang X, Lorch S, Rapaport-Kelz R, Mosher RE, Even-Shoshan O. Preoperative antibiotics and mortality in the elderly. *Annals of Surgery*. 2005; 242:107–114. [PubMed: 15973108]
24. Silber JH, Rosenbaum PR, Polsky DE, Ross RN, Even-Shoshan O, Schwartz JS, Armstrong KA, Randall TC. Do ovarian cancer outcomes differ by who gives the chemotherapy? *Journal of Clinical Oncology*. 2007; 25:1169–1175. [PubMed: 17401005]
25. Smith H. Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*. 1997; 27:325–353.

**Table 1**

Matching Variables for All Blacks, Two Nonoverlapping Matched Comparison Groups of Whites, and All Whites. Numbers are percents, except as noted.

Matching Variables at Diagnosis	Black	Matched Whites at Diagnosis	Matched Whites at Treatment	All Whites
n	806	806	806	13,756
Age (mean)	75	75	75	75
Stage 1	46	46	46	69
Stage 2	12	12	12	8
Stage 3	11	11	11	8
Stage 4	16	16	16	8
Stage Missing	14	14	14	7
Total %	100	100	100	100
Grade 1	14	14	14	32
Grade 2	21	21	22	33
Grade 3 or 4	38	38	37	24
Grade Missing	27	27	27	11
Total %	100	100	100	100
Adenocarcinoma	41	41	43	59
Endometroid	19	19	19	25
Papillary	23	23	22	10
Unknown	17	17	16	6
Total %	100	100	100	100
Diagnosis Year (mean)	1995	1995	1995	1995
Comorbidity %	43	42	43	26
Matching Variables at Treatment	Black	Matched Whites at Diagnosis	Matched Whites at Treatment	All Whites
Inadequate or No Surgery	35	18	35	14
Adequate Surgery (AS)	65	82	65	86
Total %	100	100	100	100
AS by GO Surgeon	28	28	28	24
AS by GYN Surgeon	25	43	26	50
AS by Other Surgeon	12	11	12	12
Inadequate or No Surgery	35	18	35	14
Total %	100	100	100	100
<= 50 Days to AS	47	68	49	71
> 50 Days to AS	18	13	17	15
Inadequate or No Surgery	35	18	35	14

Total %	100	100	100	100
---------	-----	-----	-----	-----