PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2012; **21**(S2): 148–154
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3260

ORIGINAL REPORT

Reweighted Mahalanobis distance matching for cluster-randomized trials with missing data

Robert A. Greevy Jr.[1,2]*, Carlos G. Grijalva[4], Christianne L. Roumie[1,3], Cole Beck[1,2], Adriana M. Hung[1,3], Harvey J. Murff[1,3], Xulei Liu[1,2] and Marie R. Griffin[1,3,4]

[1] *VA Tennessee Valley Geriatric Research Education Clinical Center (GRECC), Nashville, TN, USA*
[2] *Department of Biostatistics, Vanderbilt University, Nashville, TN, USA*
[3] *Department of Medicine, Vanderbilt University, Nashville, TN, USA*
[4] *Department of Preventive Medicine, Vanderbilt University, Nashville, TN, USA*

ABSTRACT

Purpose This paper introduces an improved tool for designing matched-pairs randomized trials. The tool allows the incorporation of clinical and other knowledge regarding the relative importance of variables used in matching and allows for multiple types of missing data. The method is illustrated in the context of a cluster-randomized trial. A Web application and an R package are introduced to implement the method and incorporate recent advances in the area.

Methods Reweighted Mahalanobis distance (RMD) matching incorporates user-specified weights and imputed values for missing data. Weight may be assigned to missingness indicators to match on missingness patterns. Three examples are presented, using real data from a cohort of 90 Veterans Health Administration sites that had at least 100 incident metformin users in 2007. Matching is utilized to balance seven factors aggregated at the site level. Covariate balance is assessed for 10 000 randomizations under each strategy: simple randomization, matched randomization using the Mahalanobis distance, and matched randomization using the RMD.

Results The RMD matching achieved better balance than simple randomization or MD randomization. In the first example, simple and MD randomization resulted in a 10% chance of seeing an absolute mean difference of greater than 26% in the percent of nonwhite patients per site; the RMD dramatically reduced that to 6%. The RMD achieved significant improvement over simple randomization even with as much as 20% of the data missing.

Conclusions Reweighted Mahalanobis distance matching provides an easy-to-use tool that incorporates user knowledge and missing data.

Received 1 September 2011; Revised 20 February 2012; Accepted 21 February 2012

INTRODUCTION

In trials designed to reflect routine care, the cluster-randomized trial is appealing. By randomly assigning interventions to physicians or hospitals instead of directly to patients, routine care settings may be studied under experimental interventions while problems such as treatment contamination can be prevented. However, the number of clusters being randomized is often relatively small. A study including thousands of patients may have randomized only a dozen hospitals. In this situation, assigning treatments with a simple randomization, for example, drawing the names of half of the hospitals out of a hat, is unacceptably risky. When the number of units being randomized is small, there is substantial risk that severe imbalance in important covariates will occur by chance.[1]

Restricted randomization methods are commonly used to reduce this risk. Cluster-randomized trials frequently have the advantage of covariate information being available on all units prior to randomization; those units are randomized all at once or in a few batches. Stratified randomization is a commonly used restricted randomization method that creates strata based on a few important covariates and then randomly assigns half of the units in each stratum to one treatment and half to the other. While providing some benefit, this approach is limited to including only a few of the important covariates and the categorization of continuous covariates into a few bins. In spite of its

*Correspondence to: Robert A. Greevy, Jr., PhD, Vanderbilt University School of Medicine, Department of Biostatistics, Nashville, TN 37232-2158, U.S.A.
E-mail: robert.greevy@vanderbilt.edu

limitations, the general concept is sound. The ideal stratification would contain exactly two similar units within each stratum. Matching prior to randomization achieves this without requiring categorization of continuous covariates, without severely limiting the number of covariates being balanced and without requiring units that match perfectly to achieve balance between the study arms.

The benefits of any restricted randomization method depend on its ability to balance important covariates, the strength of the association between the covariates and the outcome, and the study's sample size. In a non-clustered study of 132 patients who were randomly assigned treatment, Greevy *et al.* demonstrated that optimal nonbipartite matching on the Mahalanobis distance (MD) derived from 14 covariates resulted in an average increase in power equivalent to a 7% increase in sample size.[1] Moreover, this approach eliminated the rare but severe imbalances that may occur with simple randomization. Despite the method's superior performance, it presently remains less widely used than simple or stratified randomization. In cluster-randomized trials, wider adoption has been hindered by misunderstandings about matching and an absence of user-friendly tools to implement the method.

In their 2009 paper, Imai *et al.* dispelled the major misconceptions surrounding matched-pair cluster-randomization (MPCR).[2] For example, they examine the assumptions leading Martin *et al.* to recommend against MPCR in small samples.[3] When the assumption of equal cluster sizes is relaxed, as is appropriate for most practical scenarios, the MPCR that matches on cluster size and pre-treatment covariates will improve the study's efficiency and power over unmatched cluster-randomization, even with as few as six clusters. In a discussion of Imai *et al.*, Zhang and Small show the utility of optimal nonbipartite matching for achieving pre-treatment covariate balance in MPCR and for optimally selecting a set of units for study when the number of units available is greater than the number needed.[4] For observational studies utilizing matching, Rosenbaum presents a method of augmenting the distance matrix to optimally choose the number of units to study for a specified level of quality of match.[5] When the quality of matches is of greater concern than the exact number of units included, this approach can be very useful in the MPCR setting. Both approaches are incorporated into the methods presented here.

To fully realize the benefits of MPCR with several pre-treatment covariates, including continuous measures and cluster size, a multivariate distance measure is needed. To balance the cluster-specific covariate distributions, appropriate summary measures are chosen. Categorical variables may be summarized with proportions, for example, the percentage of patients taking statins. Likewise, when the shape of the distributions is not highly variable, a single summary measure may suffice for continuous covariate distributions, for example, mean low-density lipoprotein (LDL). Otherwise, multiple measures may be used, for example, the mean and standard deviation of LDL or the 10th, 50th, and 90th percentiles. Once a continuous multivariate distance measure is developed, the optimal set of matches is the set that minimizes the average distance between pairs. Lu *et al.* recently released an R package and a Web application that takes a user-created matrix of distances between units and solves for the optimal matches.[6] However, the creation of the distance matrices may create an obstacle for some researchers, and improving the utility of distance measures is an open area of research.

This paper addresses the need for a customizable distance measure that incorporates clinical and other knowledge regarding the importance of the covariates while also allowing the inclusion of covariates with missing values. The method we propose may incorporate two ways to exclude units when more units are available than can be included in the study. We introduce two user-friendly tools to implement the methodology in the form of a Web application and an R package. To aid the development of the distance measure, the Web application includes tools for assessing the quality of the matches prior to randomization and comparing them to benchmark values to assist the user in choosing covariate weights. Once the choice of weights has been finalized, the application allows the user to perform the official randomization with a user-specified random seed to allow reproducibility and, if needed, the randomization of additional study units to be added after the first set of treatment assignments has been made. The Web application and instructions on downloading the R package nbpMatching are available at http://biostat.mc.vanderbilt.edu/MatchedRandomization. Examples using real data from Veterans Health Administration (VHA) sites are presented to illustrate the method.

METHODS

Mahalanobis distance and reweighted Mahalanobis distance

The MD is a multivariate distance measure akin to the familiar Euclidean Distance; however, it has two additional benefits. First, it is scale invariant, for example, including a site's pre-treatment mean LDL in mg/dL will yield the same results as LDL in

DOI: 10.1002/pds

mmol/L. Second, it incorporates the correlations between the covariates. The effect may be thought of as down-weighting a difference in one covariate that is expected based on the differences observed in the other covariates. The MD may be written as

$$MD(x_i, x_j) = \left[(x_i - x_j)^T \boldsymbol{S}^{-1} (x_i - x_j)\right]^{1/2}$$

where x_i is the ith row of the $(n \times p)$ covariate matrix $\boldsymbol{X}$, with n subjects in the rows and p covariates in the columns, and $\boldsymbol{S}$ is the $(p \times p)$ covariance matrix of $\boldsymbol{X}$.

A limitation of the MD is that is the influence that the covariates have on the distance is driven purely by their covariance structure, not their clinical importance. The reweighted Mahalanobis distance (RMD) incorporates user-specified weights, imputed values for missing covariate data, and indicators of covariate missingness. We refer to the distance as "reweighted" to distinguish it from similarly named measures used in different settings.[7–10] The RMD may be written as

$$RMD(x_i, x_j, \boldsymbol{W}) = \left[(\tilde{x}_i - \tilde{x}_j)^T \boldsymbol{W}\tilde{\boldsymbol{S}}^{-1} \boldsymbol{W}(\tilde{x}_i - \tilde{x}_j)\right]^{1/2}$$

where $\tilde{\boldsymbol{X}}$ is the $(n \times p+q)$ covariate matrix consisting of $\boldsymbol{X}$ with the addition of indicator variables for the q covariates with missingness and missing values replaced with imputed values, $\tilde{x}_i$ is the ith row of $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{S}}$ is the $(p+q\,x\,p+q)$ covariance matrix of $\tilde{\boldsymbol{X}}$, and $\boldsymbol{W}$ is a $(p+q \times p+q)$ diagonal matrix of user-specified weights. Various methods may be used to impute the missing values, provided that the imputation is estimating an expected value without random noise added. The application presented here currently uses the R package transcan, which transforms the covariates to have their maximum correlation with the best linear combination of the other covariates and returns expected values on the original scale that may be interpreted as an expected median or mode for continuous or categorical variables, respectively.[11]

The usefulness of the imputed values will depend partially on the validity of the missing-at-random (MAR) assumption on which they are based.[12] If the MAR assumption is in question, researchers may wish to match on missingness patterns more than on the imputed missing values. This may be achieved through adjusting the weights for the missingness indicators. The Web application currently uses the same weight for all missingness indicators and a small default value of 0.1, giving preference to the MAR assumption. A weight of 0 can be used to completely eliminate the impact of the indicators.

Optimality and limitation on the number of clusters

With the use of the RMD, an $(n \times n)$ matrix of distances between units is created. Unlike a retrospective cohort study where units from one group are matched to units in another group, any of the n units may be matched to any of the other units. The optimal set of matches is the set that yields the smallest average RMD distance between the matched pairs. Thanks to advanced methods for solving this so-called optimal nonbipartite matching problem, the computational complexity no longer appreciably limits its use.[13] The Web application presented here has successfully handled up to 5000 units, well beyond the typical number of clusters in a cluster-randomized trial.

Optimally selecting a subset of clusters

In cluster-randomized studies, the cost of each hospital or experimental unit included in the study often limits the number of units that may be used because more units may be available for participation than can be included. Typically, the units that are excluded are chosen via *ad hoc* procedures. Some reasons to exclude a unit may be obvious, such as logistical difficulties unique to that unit. When there is no clear choice, the matching method can optimally select which units to drop by removing those that would create the greatest imbalance between the groups.[4] The user specifies a number of units to exclude, say k units. The application adds k units to the cohort that have the special property that they match every other unit perfectly. These units are usually referred to as "sinks" and are labeled "phantoms" by the applications presented here. Units that are matched to the phantoms are excluded from the study. In an alternate approach, the user specifies a threshold for acceptable distances, δ, that matches must meet to be included in the study. The distance matrix used for matching is augmented to select the optimal set of units satisfying the threshold.[5] This approach is equivalent to adding N units to the study that are distance δ from all other units. Any unit that is matched to one of these "near-matchers," or "chameleons," as they are called in our applications, is excluded from the study.

Evaluating performance

If examining the average difference between groups over all possible randomizations, almost any randomized method appears to balance the variables well because the expected mean difference for all of these methods is zero. However, a particular randomization may be quite poor, showing a large imbalance in the

DOI: 10.1002/pds

mean difference for a variable. Thus, balance is assessed through the 90th percentile of the absolute mean differences (AMD_90). The AMD_90 is empirically estimated via 10 000 randomizations for each strategy, and all standard errors are ≤0.1 unless specified otherwise. The simple randomization used here balances only the number of sites in each arm, drawing from a set of size $\binom{n}{n/2}$ possible randomizations for n sites. The MPCR balances the covariates of interest by restricting to a smaller set of size $2^{n/2}$ possible randomizations.

EXAMPLE STUDY

As our motivating example, we consider designing a trial to study the effects of early intensification with insulin on LDL at 12 months post-intensification. In 2006, the American Diabetes Association recommended that metformin be used as the first-line agent unless contraindicated.[14] No standard protocol currently exists for patients failing their oral anti-diabetic monotherapy shortly after starting it, that is, those who show glycosylated hemoglobin (A1c) levels of 7–9% within 3–12 months after initiation. Those failing metformin monotherapy could remain on their current treatment, intensify with insulin, or intensify to a metformin–sulfonylurea dual therapy. We consider randomly assigning VHA sites to early intensification with insulin or early intensification with dual therapy. The method proposed will be used to balance (between the two treatment arms) the distributions of covariates known to affect LDL in the VHA patient population.[15]

RESULTS

To illustrate the method, a hypothetical cohort of potential study sites is drawn from the National VHA databases, which include pharmacy, inpatient, outpatient, and laboratory records. For the VHA fiscal year 2007, 90 VHA sites serving 100–500 incident metformin users were identified. Covariate information for each patient was derived from the 365 days preceding their starting treatment; see Roumie *et al*. for covariate definitions.[15] Matching is utilized to balance seven factors aggregated at the site level: percentage of nonwhite patients, percent female, percent on statins, mean systolic blood pressure (mmHg), body mass index (BMI), A1c (%), and the number of incident metformin users (*N*). Three examples are presented. The first compares the performance of the MD and RMD to simple matching of 12 preselected sites. The second example illustrates the performance when selecting 12 sites out of the 90 potential sites via the use of phantoms or chameleons. The third example illustrates the performance when missingness is induced completely at random.

Example 1

The correlations and standard deviations for the 12 sites are shown in Table 1. The comparatively large standard deviation and low correlations for race suggest that it will be more difficult to balance than the other variables. Table 2 shows the AMD_90 for four methods: simple randomization; MD matching (MD); RMD matching with weight = 1 for race and 0 for all

Table 2. 90th percentile of the absolute mean difference (AMD_90) for four randomization methods (12 preselected sites)

Variable	Simple	MD	RMD_race	RMD_race+
Nonwhite (%)	26.8	26.1	4.8	6.4
Female (%)	1.6	1.8	1.6	2.3
On statins (%)	6.4	6.2	3.7	6.3
Systolic BP (mmHg)	1.5	1.4	2.3	2.1
BMI	1.3	0.7	1.4	0.8
A1c (%)	0.1	0.1	0.3	0.2
N patients	100.5	110.2	120.5	33.2

Methods: Simple randomization, Mahalanobis distance matching (MD), RMD matching with weight = 1 for race and weights = 0 for the other variables (RMD_race), and RMD with weights = 10 for race and BMI, 5 for statin use, and 1 otherwise (RMD_race+). 10 000 simulations run for each method. Standard error for *N* patients ≤0.3, and ≤0.1 for all other variables.

Table 1. Variable correlations and standard deviations (12 preselected sites)

	Nonwhite (%)	Female (%)	On statins (%)	Systolic BP (mmHg)	BMI	A1c (%)	*N* patients
Nonwhite	*35.2*	0.2	0.0	0.2	−0.4	0.1	0.6
Female	0.2	*2.3*	0.2	0.3	0.2	0.3	0.3
On statins	0	0.2	*6.0*	−0.2	0.2	−0.1	−0.1
Systolic BP	0.2	0.3	−0.2	*2.6*	0.4	0.7	0.1
BMI	−0.4	0.2	0.2	0.4	*1.4*	0.6	−0.6
A1c	0.1	0.3	−0.1	0.7	0.6	*0.2*	−0.4
N patients	0.6	0.3	−0.1	0.1	−0.6	−0.4	*133.4*

Standard deviations shown in italics. *N* sites = 12.

DOI: 10.1002/pds

other variables (RMD_race); and RMD with weights = 10 for race and BMI, 5 for statin use, and 1 otherwise (RMD_race+). The AMD_90 for simple randomization was 26.8%. In other words, simple randomization yielded a 10% chance of a study having a mean difference of at least 26.8% in the percent of nonwhite patients between study arms. MD showed a small improvement with 26.1%. RMD_race dramatically reduced the AMD_90 to 4.8% and RMD_race+ reduced it to 6.4%. The benefit of MD was primarily seen in BMI, reducing the AMD_90 to 0.7 from simple randomization's 1.3 and 1.4 for RMD_race. RMD_race+ had comparable balance on BMI at 0.8. Compared with the MD, RMD_race+ achieved dramatically better balance on race and cluster size at a small cost of slightly less balance on the other variables.

Example 2

Where example 1 preselected 12 from the 90 sites, MD and RMD can select an optimal subset via the use of phantoms or chameleons. The balance for sites selected via four different methods is presented in Table 3. For comparison, 500,000 sets of 12 were selected via simple random samples. On average, simple randomization performed similarly to how it did in example 1. By utilizing 78 phantoms to optimally select 12 sites, MD yielded slightly better balance on the preselected sites than RMD_race+. In this setting RMD_race+ did not balance the number of patients per site; thus, the variable "*N* patients" was also given a weight of 5, yielding RMD_race++. As in example 1, RMD_race++ outperformed MD in terms of balancing the difficult variables of race and number of patients per site, while performing almost as well as on the other variables. When selecting an optimal subset using chameleons, set with a threshold equal to the 0.2 percentile of the distance matrix, the method selected 16 sites with performance similar to the set selected via phantoms.

Table 3. 90th percentile of the absolute mean difference (AMD_90) for 12 sites selected from 90 via four methods

Variable	Simple random sample	MD with phantoms	RMD_race ++ with phantoms	RMD_race ++ with chameleons
Nonwhite (%)	23.1	3.9	1.7	2.0
Female (%)	2.1	0.9	0.9	1.1
On statins (%)	5.4	0.4	0.8	0.7
Systolic BP (mmHg)	2.1	0.5	0.7	0.8
BMI	0.7	0.1	0.1	0.1
A1c (%)	0.2	0.1	0.1	0.1
N patients	177.0	28.8	17.3	17.0

A total of 12 sites were selected via simple random samples for the simple randomization, 12 sites optimally selected using phantoms for MD and RMD_race++, and 16 sites selected using chameleons with threshold equal to the 0.2 percentile of the distance matrix. Note that a threshold equal to the 0.17 percentile yields the same 12 sites as using phantoms. RMD_race++ has weights = 10 for race and BMI, 5 for statin use and *N* patients, and 1 otherwise. A total of 500 000 simulations were performed for simple randomization and 10 000 for each of the other methods. Standard error for *N* patients ≤0.3, and ≤0.1 for all other variables.

Example 3

The performance of simple randomization will improve as the number of sites increases. When randomizing all 90 sites, simple randomization has an AMD_90 of 7.2 for race and 60.5 for patients per site. With as many as 20% of the covariate values missing, RMD_race++ outperformed simple randomization on all variables, especially race and patients per site.

DISCUSSION

The RMD provides a user-friendly method for researchers to incorporate into the matching process their clinical knowledge and the relative difficulty of balancing important covariates. Greevy *et al.* have shown that matching prior to randomization outperforms unmatched randomization in non-clustered randomized controlled trials (RCTs),[1] and Imai *et al.* have shown its benefits in clustered RCTs.[2] Zhang and Small have shown that optimal nonbipartite matching using an MD may outperform other matching methods,[4] and the current paper shows that the RMD may yield results superior to the MD when the perceived quality of the matching depends on the relative clinical importance of the variables. Moreover, the RMD may account for missing data in a sophisticated, yet highly automated, process. Table 4 shows benefits over simple randomization with up to 20% of the data missing.

Analysis for MPCR designs is a growing area of research. Recently, Imai *et al.* introduced a harmonic mean estimator for which the study inferences can be justified by the study design alone.[2] Zhang, Traskin, and Small have developed a robust test statistic for MPCR trials that outperforms linear mixed models for heavy-tailed distributions and performs nearly as well in the special case where the mixed model assumptions are true.[16] The statistic may optionally include covariate adjustment while still relying on the study design to justify inferences via the approach developed by Rosenbaum.[17] Many studies will include the covariates used in the matching as variables in the analysis model. To avoid potential bias, we discourage including the indicators for missingness that are created by RMD.[18] In addition,

DOI: 10.1002/pds

Table 4. 90th percentile of the absolute mean difference (AMD_90) for RMD_race++ with three levels of induced missingness

	Percent missing:	5%	10%	20%
Variable	Simple	RMD_race ++	RMD_race ++	RMD_race ++
Nonwhite (%)	7.2	3.2	3.8	4.9
Female (%)	0.7	0.6	0.6	0.7
On statins (%)	1.8	1.2	1.3	1.5
Systolic BP (mmHg)	0.8	0.8	0.8	0.8
BMI	0.3	0.1	0.1	0.2
A1c (%)	0.1	0.1	0.1	0.1
N patients	60.5	33.8	37.4	44.5

A total of 90 sites matched. Balance is for the true covariate values without missingness. Missingness was induced completely at random (MCAR) over all seven covariates. RMD_race++ has weights = 10 for race and BMI, 5 for statin use and *N* patients, and 1 otherwise. A total of 10 000 simulations run were performed for each method. Standard errors for *N* patients are ≤0.7, and ≤0.1 for all other variables.

the form of the covariates used in the model may also vary from the form used in the matching; for example, the model may benefit from the transformation of a covariate to account for a nonlinear association with the outcome.

In situations where the potential for imbalance was low, for example, low variability in the covariates, the benefit of up-weighting variables purely for their clinical importance was small. The choice of weights was best made with a combination of clinical knowledge and examination of the pre-randomization covariate data. Users may influence the impact of individual variables on the MD by dropping variables entirely or applying nonlinear transformations to them, for example, rank or log. When matching observational data with highly non-normal distributions, Rosenbaum recommends using a rank-based MD (MD_rank).[19] The MD_rank is equivalent to an RMD that uses a rank transformation of each variable (i.e., replacing a variable with the rank ordering of the variable and using average ranks for ties) and uses weights equal to the standard deviation of the rank-transformed variables divided by what the standard deviation of the ranks would be if there were no ties. This serves to down-weight variables with ties. Ranking is particularly useful for covariates that have outliers, and it is often desirable to down-weight a variable with numerous ties, for example, a binary variable such as "teaching hospital Y/N." The clinical importance of a particular variable may discourage researchers from down-weighting, or even to up-weight, that variable. Because the aggregated measures used in the examples have sufficient precision to prevent any ties, the MD_rank is equivalent to the MD on rank-transformed data. The effects of the rank transformation are presented in online Supplementary Tables 1 and 2.

For missing data, a user may wish to use an imputation method that is more sophisticated than the highly automated procedure used here. The Web application and the R package allow more advanced users to use a covariate matrix that incorporates their customized changes or use their own customized distance matrix. The methods applied here provide a straightforward, easily implemented method for creating optimally matched clusters for randomization in an MPCR study.

CONFLICTS OF INTEREST

The authors of this research are responsible for its content. Statements here should not be construed as endorsement by the Agency for Healthcare Research and Quality or the US Department of Health and Human Services. There were no conflicts of interest with this research.

ACKNOWLEDGEMENTS

This project was funded in part by the Agency for Healthcare Research and Quality, US Department of Health and Human Services, Contract No. HHSA2902010000161, as part of the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE 2) program. The work of R. G. was supported in part by a grant from the National Institutes of Health, P60AR056116. The work of A.M.H. was supported in full by the Career Development Program from the Department of Veterans Affairs CDA (2-031-09S) from CSR&D.

Supporting Information

Additional supporting information may be found in the online version of this article:

Supplementary Table 1: 90th percentile of the absolute mean difference (AMD_90) comparing three rank-based methods (12 preselected sites)

Supplementary Table 2: 90th percentile of the absolute mean difference (AMD _90) for 12 sites selected from 90 comparing three rank-based methods

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

1. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. *Biostatistics* Apr 2004; **5**(2):263–275.
2. Imai K, King G, Nall C. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. *Stat Sci* Feb 2009; **24**(1): 29–53.
3. Martin DC, Diehr P, Perrin EB, Koepsell TD. The effect of matching on the power of randomized community intervention studies. *Stat Med* 1993; **12**: 329–338.
4. Zhang K, Small D. Comment: the essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. *Stat Sci* Feb 2009; **24**(1): 59–64.
5. Rosenbaum PR. Optimal matching of an optimally chosen subset in observational studies. *J Comput Graph Stat* 2011; [Epub ahead of print] doi: 10.1198/jcgs.2011.09219.
6. Lu B, Greevy R, Xu X, Beck C. Optimal nonbipartite matching and its statistical applications. *Am Stat* 2011; **65**(1): 21–30.
7. Wölfel M, Ekenel H. Feature weighted Mahalanobis distance: improved robustness for Gaussian classifiers. 13th European Signal Processing Conference (EUSIPCO), Antalya, Turkey. September 2005.
8. Younis K, Karim M, Hardie R, Loomis J, Rogers S, DeSimio M. Cluster merging based on weighted Mahalanobis distance with application in digital mammograph. Proc. of IEEE Aerospace and Electronics Conference. 1998.
9. Peng J, Heistenkamp DR, Dai HK. Adaptive kernel metric nearest neighbor classification. Proc. of IEEE International Conference on Pattern Recognition. 2002.
10. Rerkrai K, Fillbrandt H. Tracking persons under partial scene occlusion using linear regression. 8th International Student Conference on Electrical Engineering. 2004.
11. Harrell FE, *et al*. Hmisc. R package version 3.9-1. January 2012. http://CRAN.R-project.org/package=Hmisc.
12. Little RJA, Rubin DB. Statistical Analysis with Missing Data (2nd edn). Wiley: New York, 2002.
13. Derigs U. Solving nonbipartite matching problems via shortest path techniques. *Ann Oper Res* 1988; **13**: 225–261.
14. Summary of revisions for the 2006 Clinical Practice Recommendations. *Diabetes Care* Jan 2006; **29**(Suppl 1): S3.
15. Roumie CL, Huizinga MM, Liu X, *et al*. The effect of incident antidiabetic regimens on lipid profiles in veterans with type 2 diabetes: a retrospective cohort. *Pharmacoepidemiol Drug Saf* 2011; **20**: 36–44.
16. Zhang K, Traskin M, Small D. A powerful and robust test statistic for randomization inference in group-randomized trials with matched pairs of groups. *Biometrics* 2011; [Epub ahead of print] doi: 10.1111/j.1541-0420.2011.01622.x.
17. Rosenbaum PR. Covariance adjustment in randomized experiments and observational studies. *Stat Sci* 2002; **17**(3): 286–327.
18. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995; **142**(12): 1255–1264.
19. Rosenbaum PR. Design of Observational Studies. Springer Series in Statistics: New York, 2010. Ch 8.