The Annals of Statistics
1992, Vol. 20, No. 2, 1079–1093

AFFINELY INVARIANT MATCHING METHODS WITH ELLIPSOIDAL DISTRIBUTIONS[1]

BY DONALD B. RUBIN AND NEAL THOMAS[2]

Harvard University and Educational Testing Service

Matched sampling is a common technique used for controlling bias in observational studies. We present a general theoretical framework for studying the performance of such matching methods. Specifically, results are obtained concerning the performance of affinely invariant matching methods with ellipsoidal distributions, which extend previous results on equal percent bias reducing methods. Additional extensions cover conditionally affinely invariant matching methods for covariates with conditionally ellipsoidal distributions. These results decompose the effects of matching into one subspace containing the best linear discriminant, and the subspace of variables uncorrelated with the discriminant. This characterization of the effects of matching provides a theoretical foundation for understanding the performance of specific methods such as matched sampling using estimated propensity scores. Calculations for such methods are given in subsequent articles.

1. Background. Matched sampling is a popular and important technique for controlling bias in observational studies. It has received increasing attention in the statistical literature in recent years [Cochran (1968); Cochran and Rubin (1973); Rubin (1973a, b), (1976a, b), (1979); Carpenter (1977); and Rosenbaum and Rubin (1983, 1985)]. The basic situation has two populations of units, treated (e.g., smokers) and control (e.g., nonsmokers), and a set of observed matching variables $\mathbf{X} = (X_1, \dots, X_p)$ (e.g., age, gender, weight). The objective is to compare the distributions of the outcome variables having adjusted for differences in the distributions of $\mathbf{X}$ in the two populations. Matched sampling is a way of adjusting for $\mathbf{X}$ through data collection.

Suppose there exist random samples from the treated and control populations of sizes N_t and N_c, respectively, where $\mathbf{X}$ is recorded on all $N_t + N_c$ units; typically, due to cost considerations, outcomes and additional covariates can only be recorded on subsamples of the initial samples. Instead of randomly choosing subsamples, often matched subsamples of sizes $n_t \le N_t$ and $n_c \le N_c$ are chosen in such a way that the distributions of $\mathbf{X}$ among the n_t and n_c

Received January 1990; revised January 1991.

[1]Work partially supported by NSF Grants SES-83-11428, SES-88-05433, SOC-76-80389 and MCS-83-01459, and National Institutes of Health Grant HD17574-01A2.

[2]Work also partially supported by the Department of Statistics at the University of Chicago, the Statistics Research consulting Group at RAND Corporation and the Research Statistics Group at Educational Testing Service.

AMS 1980 *subject classifications*. Primary 62D05, 62H05; secondary 62A05, 62H30, 62K99.

Key words and phrases. Matched sampling, bias reduction, Mahalanbois metric matching, discriminant matching, propensity score, observational studies, nonrandomized studies.

matched units are more similar than they would be in random subsamples. Commonly, $N_c \gg N_t$ and $n_c = n_t = N_t$, so that only the controls are subsampled, as when the treated group has been exposed to an unusual occurrence (e.g., radiation) and the controls are all those who were not exposed.

The standard matched-sample estimator of the treatment's effect on an outcome Y is the difference in the means of Y between the n_t and n_c matched treated and control units, $\bar{Y}_{mt} - \bar{Y}_{mc}$. If the distributions of Y differ in the treated and control populations primarily because of the treatment effect and differences in $\mathbf{X}$, then $\bar{Y}_{mt} - \bar{Y}_{mc}$ should be closer to the treatment effect than the difference of Y means in random subsamples of size n_t and n_c, $\bar{Y}_{rt} - \bar{Y}_{rc}$, thereby reflecting bias reduction arising from the matched sampling.

Here we present theoretical results on the performance of affinely invariant matching methods, defined in Section 2, with ellipsoidal distributions for $\mathbf{X}$, also defined in Section 2. The multivariate normal and the multivariate t are special cases of commonly referenced ellipsoidal distributions, and many practical matching methods are affinely invariant, such as discriminant matching, Mahalanobis metric matching, matching based on propensity scores estimated by logistic regression, and combinations of these considered in the statistics literature.

Results in Section 3 extend and generalize those in Rubin (1976a) on bias reduction in theoretically and practically important ways. In particular, we provide general theoretical results not only on bias reduction for any linear combination of $\mathbf{X}$, $Y = \alpha'\mathbf{X}$, as in Rubin (1976a), but also on the variance of $\bar{Y}_{mt} - \bar{Y}_{mc}$ relative to that of $\bar{Y}_{rt} - \bar{Y}_{rc}$, and on the expectation of the second moments of Y in the matched samples. These results are of importance to practice because they include cases in which the matching methods are based on estimated discriminant or estimated propensity scores, as well as on estimated metrics such as the Mahalanobis metric, and therefore lay the foundation for obtaining valid standard errors in samples matched using estimated propensity scores.

In Section 4, extensions are presented for matching methods that are only conditionally affinely invariant because they use a subset of the matching variables in a special way. These extensions only require conditionally ellipsoidal distributions, a generalization of the normal general location model [e.g., Olkin and Tate (1961)], and are relevant to the important practice of forcing better matching with respect to key covariates [e.g., Mahalanobis metric matching on a subset of variables within calipers of the estimated propensity scores, Rosenbaum and Rubin (1985)].

Explicit analytic expressions based on the results of Sections 3 and 4 under multivariate normality using discriminant matching are given in Rubin and Thomas (1991). Qualitative descriptions are also given indicating how these results for matching with estimated scores change for different types of nonnormal distributions; the implications of these results for practice are studied using simulation techniques and real data in Rubin and Thomas (1992).

2. Definitions and notation.

2.1. *Ellipsoidal distributions.* Ellipsoidal distributions are characterized by the fact that there exists a linear transformation of the variables that results in a spherically symmetric distribution for the transformed variables. An ellipsoidal distribution is fully specified by (i) its center of symmetry, (ii) its inner product defined by the linear transformation to sphericity and (iii) the distribution on the radii of concentric hyperspheres on which there is uniform probability density [Dempster (1969)]. Such multivariate distributions play an important role in the theory of matching methods because the symmetry allows general results to be obtained.

In the general matching situation with ellipsoidal distributions, let $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_c$ be the centers of $\mathbf{X}$ and Σ_t and Σ_c be the inner products of $\mathbf{X}$ in the treated and control populations, respectively. Although some of our results hold with weaker restrictions, we assume that $\Sigma_t \propto \Sigma_c$, so that one common linear transformation can reduce $\mathbf{X}$ to sphericity in both treated and control populations; however, the two ellipsoidal distributions may differ (e.g., normal in the control group and t_3 in the treated group).

When $\Sigma_t \propto \Sigma_c$, the ellipsoidal distributions will be called proportional. For simplicity of description, we assume finite second moments of $\mathbf{X}$ so that $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_c$ are the expectations (means) of $\mathbf{X}$ and Σ_t and Σ_c are its covariances. Little generality of practical relevance is lost by making these latter restrictions.

2.2. *Canonical form for proportional ellipsoidal distributions.* As noted by Cochran and Rubin (1973), Rubin (1976a), Efron (1975) and other authors, with proportional ellipsoidal distributions, there exists an affine transformation of $\mathbf{X}$ to the following canonical form,

$$\boldsymbol{\mu}_t = \delta \mathbf{1}, \tag{1}$$

where δ is a positive scalar constant and $\mathbf{1}$ is the unit vector,

$$\boldsymbol{\mu}_c = \mathbf{0}, \tag{2}$$

where $\mathbf{0}$ is the zero vector,

$$\Sigma_t = \sigma^2 \mathbf{I}, \tag{3}$$

where σ^2 is a positive scalar and $\mathbf{I}$ is the identity matrix and

$$\Sigma_c = \mathbf{I}. \tag{4}$$

This canonical form for the distributions of $\mathbf{X}$ is very useful because the resulting distributions are fully exchangeable (symmetric) in the coordinates of $\mathbf{X}$ so that many results can be obtained by simple symmetry arguments. For these symmetry arguments to apply with matched samples, the matching method must also possess the corresponding symmetry and so we define affinely invariant matching methods to have this property.

2.3. *Affinely invariant matching methods.* Let $(\mathscr{X}_t, \mathscr{X}_c)$ be a pair of (units by variables) data matrices, where $\mathscr{X}_t$ is the N_t by p matrix with elements (X_{tij}) and $\mathscr{X}_c$ is the N_c by p matrix with elements (X_{cij}). A general matching method is a mapping from $(\mathscr{X}_t, \mathscr{X}_c)$ to a pair of sets of indices, (T, C), of those units chosen in the matched samples, where T has n_t elements from $(1, \dots, N_t)$ and C has n_c elements from $(1, \dots, N_c)$. (A more restrictive definition of general matching would require $n_t = n_c$, with a $1:1$ correspondence between the elements of T and C). An affinely invariant matching method is one such that the matching output is the same following any affine transformation, A, of $\mathbf{X}$:

$$(\mathscr{X}_t, \mathscr{X}_c) \mapsto (T, C),$$

implies

$$(A(\mathscr{X}_t), A(\mathscr{X}_c)) \mapsto (T, C).$$

Matching methods based on population or sample inner products, such as discriminant matching or Mahalanobis metric matching, are affinely invariant, as are common methods using propensity scores based on linear logistic regression estimators, which are affinely invariant as noted in Efron (1975). Methods that are not affinely invariant include those where one coordinate of $\mathbf{X}$ is treated differently from the other components (e.g., weighted in the Mahalanobis metric to reflect greater importance) or where nonlinear estimators (e.g., of the discriminant) are used, such as trimmed moment estimators.

When an affinely invariant matching method is used with proportional ellipsoidal distributions, the canonical form for the distributions of $\mathbf{X}$ given in (1)–(4) can be assumed. This canonical form will be used throughout the remainder of Section 2 and in Section 3.

2.4. *The best linear discriminant.* A particularly important linear combination of $\mathbf{X}$ is the best linear discriminant, $(\boldsymbol{\mu}_t - \boldsymbol{\mu}_c)'\Sigma_c^{-1}\mathbf{X}$, which is proportional to $\mathbf{1}'\mathbf{X}$ in the canonical form of the distributions defined by (1)–(4). The standardized discriminant, Z, is defined as the discriminant with unit variance in the control population,

$$Z = \mathbf{1}'\mathbf{X}/\sqrt{p}\,. \tag{5}$$

The standardized discriminant is a log-likelihood ratio statistic when $\mathbf{X}_t$ and $\mathbf{X}_c$ are multivariate normal with proportional covariance matrices, and it has the maximum difference in population means among all standardized linear combinations in the more general setting.

An arbitrary linear combination $Y = \boldsymbol{\alpha}'\mathbf{X}$ can be expressed as a sum of components along and orthogonal to the standardized discriminant; for notational simplicity, let Y be standardized, $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. Let W be the standardized linear combination of $\mathbf{X}$ orthogonal to Z,

$$W = \boldsymbol{\gamma}'\mathbf{X}, \qquad \boldsymbol{\gamma}'\mathbf{1} = 0, \qquad \boldsymbol{\gamma}'\boldsymbol{\gamma} = 1, \tag{6}$$

with γ chosen so that

$$Y = \rho Z + \sqrt{1 - \rho^2}\, W, \tag{7}$$

where ρ is the correlation of Y with the standardized discriminant Z. By their construction, Z and W have several properties useful in the derivations. In the control population,

$$\begin{gathered} E(\bar{Z}_{rc}) = E(\bar{W}_{rc}) = E(\bar{Y}_{rc}) = 0, \\ \operatorname{var}(\bar{Z}_{rc}) = \operatorname{var}(\bar{W}_{rc}) = \operatorname{var}(\bar{Y}_{rc}) = \frac{1}{n_c}, \end{gathered} \tag{8}$$

where the subscript rc refers to a randomly chosen sample of n_c control units. Likewise, in the treated population,

$$\begin{gathered} E(\bar{Z}_{rt}) = \delta\sqrt{p}\,, \qquad E(\bar{W}_{rt}) = 0, \qquad E(\bar{Y}_{rt}) = \rho\delta\sqrt{p}\,, \\ \operatorname{var}(\bar{Z}_{rt}) = \operatorname{var}(\bar{W}_{rt}) = \operatorname{var}(\bar{Y}_{rt}) = \frac{\sigma^2}{n_t}, \end{gathered} \tag{9}$$

where the subscript rt refers to a randomly chosen sample of n_t treated units.

Corollaries in Section 3 decompose the effects of matching into two pieces: the effect on the best linear discriminant Z and the effect on any covariate W orthogonal to Z.

3. Decomposing the matching effects.

3.1. *Results based on symmetry.*

THEOREM 3.1. *Suppose an affinely invariant matching method is applied with fixed sample sizes (N_t, N_c, n_t, n_c) and proportional ellipsoidal distributions, which are represented in canonical form. Then,*

$$E(\bar{\mathbf{X}}_{mt}) \propto \mathbf{1}, \tag{10}$$

$$E(\bar{\mathbf{X}}_{mc}) \propto \mathbf{1} \tag{11}$$

and

$$\operatorname{var}(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}) \propto \mathbf{I} + c\mathbf{1}\mathbf{1}', \qquad c > -1/p, \tag{12}$$

where $\bar{\mathbf{X}}_{mt}$ and $\bar{\mathbf{X}}_{mc}$ are the mean vectors in the matched treated and control samples and $E(\cdot)$ and $\operatorname{var}(\cdot)$ are the expectation and variance over repeated random draws from the initial populations of treated and control units. Furthermore,

$$E(v_{mt}(\mathbf{X})) \propto \mathbf{I} + c_t\mathbf{1}\mathbf{1}', \qquad c_t \geq -1/p, \tag{13}$$

$$E(v_{mc}(\mathbf{X})) \propto \mathbf{I} + c_c\mathbf{1}\mathbf{1}', \qquad c_c \geq -1/p, \tag{14}$$

where $v_{mt}(\mathbf{X})$ and $v_{mc}(\mathbf{X})$ are the sample covariance matrices of $\mathbf{X}$ in the *matched treated and control samples with divisors* $(n_t - 1)$ *and* $(n_c - 1)$. *If* $\delta = 0$, *then*

$$E(\overline{\mathbf{X}}_{mt}) = E(\overline{\mathbf{X}}_{mc}) = \mathbf{0},$$
$$\mathrm{var}(\overline{\mathbf{X}}_{mt} - \overline{\mathbf{X}}_{mc}) \propto \mathbf{I},$$
$$E(v_{mt}(\mathbf{X})) \propto \mathbf{I},$$
$$E(v_{mc}(\mathbf{X})) \propto \mathbf{I}.$$

PROOF OF THEOREM 3.1. The proof of Theorem 3.1 follows from symmetry with little derivation. The expectations of the sample means of each coordinate of $\mathbf{X}$ must be the same and thus proportional to $\mathbf{1}$ in both treated and matched control samples. The covariance matrices must be exchangeable in the treated and matched control samples. The general form for the covariance matrix of exchangeable variables is proportional to

$$\mathbf{I} + c\mathbf{1}\mathbf{1}', \qquad c \geq -1/p.$$

When $\delta = 0$, the complete rotational symmetry implies the final set of claims. □

The symmetry results of Theorem 3.1 imply that any W orthogonal to the discriminant Z has the same distribution, as summarized in the following corollary.

COROLLARY 3.1. *The quantities* $\mathrm{var}(\overline{W}_{mt} - \overline{W}_{mc})$, $E(v_{mt}(W))$ *and* $E(v_{mc}(W))$ *take the same three values for all standardized* W *orthogonal to* Z. *The analogous three results apply for statistics in random subsamples indexed by* rt *and* rc. *Since* Z *is the discriminant, defined without regard to the choice of* Y, *the analogous quantities for* Z *are also the same for all* Y.

3.2. *Corollaries that decompose the effect of matching.* Although Theorem 3.1 and Corollary 3.1 follow almost immediately from the symmetry of ellipsoidal distributions and restrictions placed on the matching algorithms, their consequences are not as apparent. The corollaries that follow from it, stated here and proved in Section 3.3, show that under the conditions of Theorem 3.1, the moments of any Y in the matched samples are determined by the moments in the matched samples of the discriminant and any single covariate uncorrelated with the discriminant, and furthermore, that those moments of Y involve Y only through ρ.

Suppose proportional ellipsoidal distributions with $\boldsymbol{\mu}_t \neq \boldsymbol{\mu}_c$, and affinely invariant matching methods, and let Y have correlation ρ with the best linear discriminant Z and correlation $\sqrt{1-\rho^2}$ with W uncorrelated with Z. Then, the following four corollaries hold.

COROLLARY 3.2. *Matching is equal percent bias reducing, EPBR (Rubin, 1976a),*

$$\frac{E(\overline{Y}_{mt} - \overline{Y}_{mc})}{E(\overline{Y}_{rt} - \overline{Y}_{rc})} = \frac{E(\overline{Z}_{mt} - \overline{Z}_{mc})}{E(\overline{Z}_{rt} - \overline{Z}_{rc})}, \tag{15}$$

where the subscript rt refers to a randomly chosen sample of n_t treated units, and the subscript rc refers to a randomly chosen sample of n_c control units. Equation (15) implies that the percent reduction in bias is the same for any linear combination of $\mathbf{X}$, *because* $E(\bar{Z}_{mt} - \bar{Z}_{mc})/E(\bar{Z}_{rt} - \bar{Z}_{rc})$ *takes the same value for all Y.*

COROLLARY 3.3. *The matching is ρ^2-proportionate modifying of the variance of the difference in matched sample means,*

$$\frac{\operatorname{var}(\bar{Y}_{mt} - \bar{Y}_{mc})}{\operatorname{var}(\bar{Y}_{rt} - \bar{Y}_{rc})} = \rho^2 \frac{\operatorname{var}(\bar{Z}_{mt} - \bar{Z}_{mc})}{\operatorname{var}(\bar{Z}_{rt} - \bar{Z}_{rc})} + (1 - \rho^2) \frac{\operatorname{var}(\bar{W}_{mt} - \bar{W}_{mc})}{\operatorname{var}(\bar{W}_{rt} - \bar{W}_{rc})}, \tag{16}$$

where the ratios

$$\frac{\operatorname{var}(\bar{Z}_{mt} - \bar{Z}_{mc})}{\operatorname{var}(\bar{Z}_{rt} - \bar{Z}_{rc})}, \quad \frac{\operatorname{var}(\bar{W}_{mt} - \bar{W}_{mc})}{\operatorname{var}(\bar{W}_{rt} - \bar{W}_{rc})},$$

take the same two values for all Y.

COROLLARY 3.4. *The matching is ρ^2-proportionate modifying of the expectations of the sample variances*

$$\frac{E(v_{mt}(Y))}{E(v_{rt}(Y))} = \rho^2 \frac{E(v_{mt}(Z))}{E(v_{rt}(Z))} + (1 - \rho^2) \frac{E(v_{mt}(W))}{E(v_{rt}(W))} \tag{17}$$

and

$$\frac{E(v_{mc}(Y))}{E(v_{rc}(Y))} = \rho^2 \frac{E(v_{mc}(Z))}{E(v_{rc}(Z))} + (1 - \rho^2) \frac{E(v_{mc}(W))}{E(v_{rc}(W))}, \tag{18}$$

where $v_{rt}(\cdot)$ is the sample variance of n_t randomly chosen treated units (computed using $n_t - 1$ in the denominator) and likewise for $v_{rc}(\cdot)$ and n_c randomly chosen control units (using $n_c - 1$). The ratios

$$\frac{E(v_{mt}(Z))}{E(v_{rt}(Z))}, \quad \frac{E(v_{mt}(W))}{E(v_{rt}(W))}, \quad \frac{E(v_{mc}(Z))}{E(v_{rc}(Z))}, \quad \frac{E(v_{mc}(W))}{E(v_{rc}(W))},$$

take the same four values for all Y.

COROLLARY 3.5. *When* $\boldsymbol{\mu}_t = \boldsymbol{\mu}_c$,

$$E(\bar{Y}_{mt} - \bar{Y}_{mc}) = E(\bar{Y}_{rt} - \bar{Y}_{rc}) = 0,$$

and the ratios

$$\frac{\operatorname{var}(\bar{Y}_{mt} - \bar{Y}_{mc})}{\operatorname{var}(\bar{Y}_{rt} - \bar{Y}_{rc})}, \quad \frac{E(v_{mt}(Y))}{E(v_{rt}(Y))}, \quad \frac{E(v_{mc}(Y))}{E(v_{rc}(Y))},$$

take the same three values for all Y.

3.3. *Proofs of the corollaries.*

PROOF OF COROLLARY 3.2. After matching, from (7),

$$E(\bar{Y}_{mt} - \bar{Y}_{mc}) = \rho E(\bar{Z}_{mt} - \bar{Z}_{mc}) + \left(\sqrt{1-\rho^2}\right) E(\bar{W}_{mt} - \bar{W}_{mc}),$$

where by (6),

$$E(\bar{W}_{mt} - \bar{W}_{mc}) = \boldsymbol{\gamma}' E(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}).$$

But by Theorem 3.1, $E(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}) \propto \mathbf{1}$, and from (6), $\boldsymbol{\gamma}'\mathbf{1} = 0$. Hence,

$$E(\bar{Y}_{mt} - \bar{Y}_{mc}) = \rho E(\bar{Z}_{mt} - \bar{Z}_{mc}), \tag{19}$$

and Corollary 3.2 follows by noting from (7) that $E(\bar{Y}_{rt} - \bar{Y}_{rc}) = \rho E(\bar{Z}_{rt} - \bar{Z}_{rc})$. □

PROOF OF COROLLARY 3.3. After matching, from (7),

$$\operatorname{var}(\bar{Y}_{mt} - \bar{Y}_{mc}) = \rho^2 \operatorname{var}(\bar{Z}_{mt} - \bar{Z}_{mc}) + (1-\rho^2)\operatorname{var}(\bar{W}_{mt} - \bar{W}_{mc}),$$

because from (5) and (6),

$$\operatorname{cov}(\bar{Z}_{mt} - \bar{Z}_{mc}, \bar{W}_{mt} - \bar{W}_{mc}) = \frac{1}{\sqrt{p}} \mathbf{1}' \operatorname{var}(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc})\boldsymbol{\gamma},$$

which from (12), is proportional to

$$\mathbf{1}'(\mathbf{I} + c\mathbf{1}\mathbf{1}')\boldsymbol{\gamma} = \mathbf{1}'\boldsymbol{\gamma} + cp\mathbf{1}'\boldsymbol{\gamma} = 0.$$

Equation (16) follows because in random subsamples, the treated and control subsamples are independent with

$$\operatorname{var}(\bar{Y}_{rt} - \bar{Y}_{rc}) = \operatorname{var}(\bar{Z}_{rt} - \bar{Z}_{rc}) = \operatorname{var}(\bar{W}_{rt} - \bar{W}_{rc}) = \frac{\sigma^2}{n_t} + \frac{1}{n_c}.$$

The final claim of Corollary 3.3 follows from Corollary 3.1. □

PROOF OF COROLLARY 3.4. After matching, in the treated sample,

$$E(v_{mt}(Y)) = \rho^2 E(v_{mt}(Z)) + (1-\rho^2) E(v_{mt}(W))$$

because the expected matched treated sample covariance of Z and W is

$$\frac{1}{\sqrt{p}} E(\mathbf{1}' v_{mt}(\mathbf{X})\boldsymbol{\gamma}) \propto \mathbf{1}'(\mathbf{I} + c_t p \mathbf{1}\mathbf{1}')\boldsymbol{\gamma} = 0.$$

Noting that $E(v_{rt}(Y)) = E(v_{rt}(Z)) = E(v_{rt}(W)) = 1$, implies equation (17), and an analogous derivation for the control sample establishes (18). Corollary 3.1 completes the proof of Corollary 3.4. □

Corollary 3.5 follows immediately from symmetry considerations.

4. Extensions involving a special set of covariates.

4.1. *Conditionally affinely invariant matching methods.* An important class of extended results covers methods that treat a subset of the covariates, denoted by $\mathbf{X}^{(s)} = (X_1, \dots, X_s)$, differently from the remaining covariates, $\mathbf{X}^{(r)} = (X_{s+1}, \dots, X_p)$. For example, in a study of the effects of smoking on health in human populations, a match for each smoker might be selected according to the criteria (a) the closest nonsmoker with respect to the discriminant (computed using numerous personal characteristics) who exactly matches the smoker's gender and is within ± 5 years of the smoker's age or (b) the closest nonsmoker with respect to the Mahalanobis metric on age and an index measuring environmental exposure to carcinogens who is also within $\pm 1/4$ standard deviations of the smoker on the discriminant, as in Rosenbaum and Rubin (1985). In these examples, the special matching variables, $\mathbf{X}^{(s)}$ denote (a) gender and age and (b) age and environmental exposure.

Using a construction similar to the definition of affinely invariant matching in Section 2.3, these matching methods are called *conditionally affinely invariant* because they satisfy the condition that their matching output is the same following an affine transformation A of $\mathbf{X}^{(r)}$:

$$\left((\mathscr{X}_t^{(s)}, \mathscr{X}_t^{(r)}), (\mathscr{X}_c^{(s)}, \mathscr{X}_c^{(r)})\right) \mapsto (T, C)$$

implies

$$\left((\mathscr{X}_t^{(s)}, A\mathscr{X}_t^{(r)}), (\mathscr{X}_c^{(s)}, A\mathscr{X}_c^{(r)})\right) \mapsto (T, C).$$

Conditionally affinely invariant matching methods include affinely invariant matching methods as a subclass. When using conditionally affinely invariant matching methods, similar, but weaker results are obtained under correspondingly weaker distributional assumptions that remove the requirement of rotational invariance for the special variables $\mathbf{X}^{(s)}$.

4.2. *Conditionally ellipsoidal distributions.* Distributions satisfying this weaker set of restrictions are called *conditionally ellipsoidal distributions*: The conditional distribution of $\mathbf{X}^{(r)}$ given $\mathbf{X}^{(s)}$ is ellipsoidal with the conditional mean a linear function of $\mathbf{X}^{(s)}$ and constant conditional covariance matrix. A special case of conditionally ellipsoidal distributions is the normal general location model widely discussed in the statistics literature, starting with Olkin and Tate (1961).

Denote the covariance matrices and means for the two subsets of covariates by $\Sigma_t^{(s)}$, $\Sigma_t^{(r)}$, $\boldsymbol{\mu}_t^{(s)}$ and $\boldsymbol{\mu}_t^{(r)}$ in the treated population, and $\Sigma_c^{(s)}$, $\Sigma_c^{(r)}$, $\boldsymbol{\mu}_c^{(s)}$ and $\boldsymbol{\mu}_c^{(r)}$ in the control population, and the conditional covariance matrices and means of $\mathbf{X}^{(r)}$ given $\mathbf{X}^{(s)}$ by $\Sigma_t^{(r|s)}$, $\Sigma_c^{(r|s)}$, $\boldsymbol{\mu}_t^{(r|s)}$ and $\boldsymbol{\mu}_c^{(r|s)}$. The distributions are called proportional conditionally ellipsoidal distributions if the conditional distributions $\mathbf{X}^{(r)}|\mathbf{X}^{(s)}$ are ellipsoidal in both the treated and control populations with proportional conditional covariance matrices, $\Sigma_t^{(r|s)} \propto \Sigma_c^{(r|s)}$ and common linear regressions of the $(p-s)$ covariates in $\mathbf{X}^{(r)}$ on the s variables in $\mathbf{X}^{(s)}$, $B = (\boldsymbol{\beta}_{s+1}, \dots, \boldsymbol{\beta}_p)$. The elements in the kth column of the s by $(p-s)$

matrix, $\mathbf{B}$, are the multiple regression coefficients of X_k on $(X_1, \dots, X_s)$ for $k = (s+1), \dots, p$.

The special case of proportional conditionally ellipsoidal distributions with $\mathbf{X}^{(s)}$ binomial or multinomial and $\mathbf{X}^{(r)}$ multivariate normal has been studied extensively in the discrimination literature [e.g., Krzanowski (1975), (1980); Daudin (1986)]. This model gives rise to a linear logistic regression model for predicting population membership based on the covariates, $\mathbf{X}$, and thus is relevant to the applied practice that estimates linear propensity scores using logistic regression [Rosenbaum and Rubin (1985)].

4.3. *Canonical form for conditionally ellipsoidal distributions.* Using conditionally affinely invariant matching methods with proportional conditionally ellipsoidal distributions, a canonical form for the distributions can be assumed corresponding to the canonical form used in Section 2.2,

$$\boldsymbol{\mu}_t^{(r)} \propto \mathbf{1}, \tag{20}$$

$$\boldsymbol{\mu}_c^{(r)} = \mathbf{0}, \tag{21}$$

$$\Sigma_t = \begin{bmatrix} \Sigma_t^{(s)} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix} \tag{22}$$

and

$$\Sigma_c = \begin{bmatrix} \Sigma_c^{(s)} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{23}$$

This form is obtained by leaving the $\mathbf{X}^{(s)}$ unchanged and letting the canonical $\mathbf{X}^{(r)}$ be defined as the components of $\mathbf{X}^{(r)}$ uncorrelated with $\mathbf{X}^{(s)}$: $\mathbf{X}^{(r)} - \mathbf{B}\mathbf{X}^{(s)}$. When the covariates are transformed to this canonical form, the conditional distributions of $\mathbf{X}^{(r)}|\mathbf{X}^{(s)}$ satisfy

$$\boldsymbol{\mu}_t^{(r|s)} = \boldsymbol{\mu}_t^{(r)} \propto \mathbf{1}, \tag{24}$$

$$\boldsymbol{\mu}_c^{(r|s)} = \boldsymbol{\mu}_c^{(r)} = \mathbf{0} \tag{25}$$

and

$$\Sigma_t^{(r|s)} = \Sigma_t^{(r)} \propto \mathbf{I}, \tag{26}$$

$$\Sigma_c^{(r|s)} = \Sigma_c^{(r)} \propto \mathbf{I}, \tag{27}$$

so that the distributions of $(\mathbf{X}^{(s)}, \mathbf{X}^{(r)})$ are exchangeable under permutations of components of $\mathbf{X}^{(r)}$ conditional on $\mathbf{X}^{(s)}$ in both populations, and thus the unconditional distributions of $(\mathbf{X}^{(s)}, \mathbf{X}^{(r)})$ are also exchangeable under permutations of components of $\mathbf{X}^{(r)}$.

4.4. *Representation for a linear combination Y.* An arbitrary standardized linear combination Y can be represented in a simple form analogous to (7),

$$Y = \rho\mathscr{Z} + \left(\sqrt{1-\rho^2}\right)\mathscr{W}, \tag{28}$$

where the vectors $\mathscr{Y}$ and $\mathscr{W}$ are the standardized projections of Y along and orthogonal to the subspace $\{\mathbf{X}^{(s)}, Z\}$, where Z requires definition. It is the standardized discriminant of the covariates uncorrelated with $\mathbf{X}^{(s)}$,

$$\left[\left(\Sigma_c^{(r|s)}\right)^{-1}\left(\boldsymbol{\mu}_t^{(r)} - \boldsymbol{\mu}_c^{(r)} - \mathbf{B}'\left(\boldsymbol{\mu}_t^{(s)} - \boldsymbol{\mu}_c^{(s)}\right)\right)\right]'\mathbf{X}^{(r)}, \tag{29}$$

or in canonical form,

$$Z = \mathbf{1}'\mathbf{X}^{(r)}/\sqrt{p-s}\,, \tag{30}$$

except when $\boldsymbol{\mu}_t^{(r)} = \boldsymbol{\mu}_c^{(r)}$, in which case Z is defined to be 0. Writing $\mathscr{Y}$ and $\mathscr{W}$ as

$$\mathscr{Y} = \boldsymbol{\psi}'\mathbf{X} = \left(\boldsymbol{\psi}^{(s)\prime}, \boldsymbol{\psi}^{(r)\prime}\right)\begin{pmatrix}\mathbf{X}^{(s)}\\ \mathbf{X}^{(r)}\end{pmatrix}, \tag{31}$$

$$\mathscr{W} = \boldsymbol{\gamma}'\mathbf{X} = \left(\boldsymbol{\gamma}^{(s)\prime}, \boldsymbol{\gamma}^{(r)\prime}\right)\begin{pmatrix}\mathbf{X}^{(s)}\\ \mathbf{X}^{(r)}\end{pmatrix}, \tag{32}$$

a characterization of $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$, in canonical form, paralleling (6), is summarized in the following lemma.

LEMMA 4.1. *The coefficients* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\psi}$ *must satisfy*

$$\boldsymbol{\gamma}^{(s)} = \mathbf{0} \tag{33}$$

and

$$\boldsymbol{\gamma}^{(r)\prime}\boldsymbol{\psi}^{(r)} = 0. \tag{34}$$

When $\boldsymbol{\mu}_t^{(r)} \neq \boldsymbol{\mu}_c^{(r)}$,

$$\boldsymbol{\psi}^{(r)} \propto \mathbf{1}, \tag{35}$$

$$\mathbf{1}'\boldsymbol{\gamma}^{(r)} = 0. \tag{36}$$

When $\boldsymbol{\mu}_t^{(r)} = \boldsymbol{\mu}_c^{(r)}$,

$$\boldsymbol{\psi}^{(r)} = \mathbf{0}. \tag{37}$$

PROOF OF LEMMA 4.1. First, (33) is true because $\mathscr{W}$ is a linear combination of the X_i uncorrelated with $\{\mathbf{X}^{(s)}, Z\}$ and thus with $\{\mathbf{X}^{(s)}\}$; (34) follows from (33) and the fact that $\mathscr{Y}$ is uncorrelated with $\mathscr{W}$; (35) and (37) follow from the definition of Z in (30), and (34) and (35) imply (36). □

4.5. *Extensions of results in Section* 3.3.

THEOREM 4.1. *Suppose a conditionally affinely invariant matching method is applied with fixed sample sizes* (N_t, N_c, n_t, n_c) *and proportional conditionally ellipsoidal distributions, which are represented in canonical form. Then,*

$$E\left(\overline{\mathbf{X}}_{mt}\right) = \begin{bmatrix} E\left(\overline{\mathbf{X}}_{mt}^{(s)}\right) \\ \mu_{mt}^{(r)}\mathbf{1}\end{bmatrix}, \tag{38}$$

$$E\left(\overline{\mathbf{X}}_{mc}\right) = \begin{bmatrix} E\left(\overline{\mathbf{X}}_{mc}^{(s)}\right) \\ \mu_{mc}^{(r)}\mathbf{1}\end{bmatrix}, \tag{39}$$

where $\mu_{mt}^{(r)}$ *is the common mean of each component of* $\overline{\mathbf{X}}_{mt}^{(r)}$ *and likewise for* $\mu_{mc}^{(r)}$ *and* $\overline{\mathbf{X}}_{mc}^{(r)}$. *The variance of the difference in matched samples means is given by*

$$(40)\qquad \operatorname{var}\left(\overline{\mathbf{X}}_{mt} - \overline{\mathbf{X}}_{mc}\right) = \begin{bmatrix} \operatorname{var}\left(\overline{\mathbf{X}}_{mt}^{(s)} - \overline{\mathbf{X}}_{mc}^{(s)}\right) & \mathbf{C}\mathbf{1}' \\ \mathbf{1}\mathbf{C}' & k(\mathbf{I} + c_0\mathbf{1}\mathbf{1}') \end{bmatrix},$$

where $k \geq 0$, $c_0 \geq -1/(p-s)$, *and* $\mathbf{C}' = (c_1, \dots, c_s)$. *Furthermore,*

$$(41)\qquad E\left(v_{mt}(\mathbf{X})\right) = \begin{bmatrix} E\left(v_{mt}(\mathbf{X}^{(s)})\right) & \mathbf{C}_t\mathbf{1}' \\ \mathbf{1}\mathbf{C}_t' & k_t(\mathbf{I} + c_{t0}\mathbf{1}\mathbf{1}') \end{bmatrix},$$

$$(42)\qquad E\left(v_{mc}(\mathbf{X})\right) = \begin{bmatrix} E\left(v_{mc}(\mathbf{X}^{(s)})\right) & \mathbf{C}_c\mathbf{1}' \\ \mathbf{1}\mathbf{C}_c' & k_c(\mathbf{I} + c_{c0}\mathbf{1}\mathbf{1}') \end{bmatrix},$$

where $k_t \geq 0$, $c_{t0} \geq -1/(p-s)$, $\mathbf{C}_t' = (c_{t1}, \dots, c_{ts})$, *and analogously for the matched control sample covariance matrix. When* $\boldsymbol{\mu}_t^{(r)} = \boldsymbol{\mu}_c^{(r)}$,

$$(43)\qquad E\left(\overline{\mathbf{X}}_{mt}\right) = \begin{bmatrix} E\left(\overline{\mathbf{X}}_{mt}^{(s)}\right) \\ \mathbf{0} \end{bmatrix},$$

$$(44)\qquad E\left(\overline{\mathbf{X}}_{mc}\right) = \begin{bmatrix} E\left(\overline{\mathbf{X}}_{mc}^{(s)}\right) \\ \mathbf{0} \end{bmatrix},$$

$$(45)\qquad \operatorname{var}\left(\overline{\mathbf{X}}_{mt}^{(r)} - \overline{\mathbf{X}}_{mc}^{(r)}\right) = \begin{bmatrix} \operatorname{var}\left(\overline{\mathbf{X}}_{mt}^{(s)} - \overline{\mathbf{X}}_{mc}^{(s)}\right) & \mathbf{0} \\ \mathbf{0} & k\mathbf{I} \end{bmatrix},$$

$$(46)\qquad E\left(v_{mt}(\mathbf{X}^{(r)})\right) = \begin{bmatrix} E\left(v_{mt}(\mathbf{X}^{(s)})\right) & \mathbf{0} \\ \mathbf{0} & k_t\mathbf{I} \end{bmatrix},$$

$$(47)\qquad E\left(v_{mc}(\mathbf{X}^{(r)})\right) = \begin{bmatrix} E\left(v_{mc}(\mathbf{X}^{(s)})\right) & \mathbf{0} \\ \mathbf{0} & k_c\mathbf{I} \end{bmatrix}.$$

The proof of Theorem 4.1 is nearly identical to the proof of Theorem 3.1. The only difference is that covariances between components in $\mathbf{X}^{(s)}$ and $\mathbf{X}^{(r)}$ appear in Theorem 4.1, which have no analog in Theorem 3.1; these covariances are exchangeable in the coordinates of $\mathbf{X}^{(r)}$. Under the additional condition $\boldsymbol{\mu}_t^{(r)} = \boldsymbol{\mu}_c^{(r)}$, the rotational invariance of $\mathbf{X}^{(r)}$ implies the further simplifications in (43)–(47). The symmetry results of Theorem 4.1 imply that any $\mathscr{W}$ orthogonal to $\mathscr{G}$ has the same distribution, as is summarized in the following corollary.

COROLLARY 4.1. *The quantities* $\operatorname{var}(\overline{\mathscr{W}}_{mt} - \overline{\mathscr{W}}_{mc})$, $E(v_{mt}(\mathscr{W}))$ *and* $E(v_{mc}(\mathscr{W}))$ *take the same three values for all standardized Y. The analogous results apply for statistics of random subsamples indexed by rt and rc. However, the corresponding expressions involving* $\mathscr{G}$ *do depend on the choice of Y.*

Under the conditions of Theorem 4.1, using the linear combination, Y, defined in (28), the following three corollaries hold.

COROLLARY 4.2. *The percent bias reduction of Y equals the percent bias reduction of Y in the subspace $\{\mathbf{X}^{(s)}, Z\}$,*

$$\frac{E\left(\overline{Y}_{mt} - \overline{Y}_{mc}\right)}{E\left(\overline{Y}_{rt} - \overline{Y}_{rc}\right)} = \frac{E\left(\overline{\mathscr{Y}}_{mt} - \overline{\mathscr{Y}}_{mc}\right)}{E\left(\overline{\mathscr{Y}}_{rt} - \overline{\mathscr{Y}}_{rc}\right)}. \tag{48}$$

COROLLARY 4.3. *The matching is ρ^2-proportionate modifying of the variance of the difference in matched sample means,*

$$\begin{aligned} &\frac{\operatorname{var}\left(\overline{Y}_{mt} - \overline{Y}_{mc}\right)}{\operatorname{var}\left(\overline{Y}_{rt} - \overline{Y}_{rc}\right)} \\ &\quad = \rho^2 \frac{\operatorname{var}\left(\overline{\mathscr{Y}}_{mt} - \overline{\mathscr{Y}}_{mc}\right)}{\operatorname{var}\left(\overline{\mathscr{Y}}_{rt} - \overline{\mathscr{Y}}_{rc}\right)} + (1 - \rho^2) \frac{\operatorname{var}\left(\overline{\mathscr{W}}_{mt} - \overline{\mathscr{W}}_{mc}\right)}{\operatorname{var}\left(\overline{\mathscr{W}}_{rt} - \overline{\mathscr{W}}_{rc}\right)}, \end{aligned} \tag{49}$$

where the ratio $\operatorname{var}(\overline{\mathscr{W}}_{mt} - \overline{\mathscr{W}}_{mc})/\operatorname{var}(\overline{\mathscr{W}}_{rt} - \overline{\mathscr{W}}_{rc})$ *takes the same value for all Y.*

COROLLARY 4.4. *The matching is ρ^2-proportionate modifying of the expectations of the sample variances,*

$$\frac{E(v_{mt}(Y))}{E(v_{rt}(Y))} = \rho^2 \frac{E(v_{mt}(\mathscr{Y}))}{E(v_{rt}(\mathscr{Y}))} + (1 - \rho^2) \frac{E(v_{mt}(\mathscr{W}))}{E(v_{rt}(\mathscr{W}))} \tag{50}$$

and

$$\frac{E(v_{mc}(Y))}{E(v_{rc}(Y))} = \rho^2 \frac{E(v_{mc}(\mathscr{Y}))}{E(v_{rc}(\mathscr{Y}))} + (1 - \rho^2) \frac{E(v_{mc}(\mathscr{W}))}{E(v_{rc}(\mathscr{W}))}, \tag{51}$$

where the ratios $E(v_{mt}(\mathscr{W}))/E(v_{rt}(\mathscr{W}))$ and $E(v_{mc}(\mathscr{W}))/E(v_{rc}(\mathscr{W}))$ take the same two values for all Y.

A result corresponding to Corollary 3.5 with $\boldsymbol{\mu}_t = \boldsymbol{\mu}_c$ can be obtained, but the simplification that occurs in the case when there are no special covariates is not present for the extended results.

PROOF OF COROLLARY 4.2. The proof is similar to that of Corollary 3.2 after noting from (32) and (33) that $\mathscr{W} = \boldsymbol{\gamma}^{(r)\prime}\mathbf{X}^{(r)}$ and from (36), (38) and (33) that $\boldsymbol{\gamma}^{(r)\prime}E(\overline{\mathbf{X}}^{(r)}_{mt} - \overline{\mathbf{X}}^{(r)}_{mc}) = 0$. □

PROOF OF COROLLARY 4.3. The proof is analogous to that of Corollary 3.3; from Theorem 4.1 and (33),

$$\operatorname{cov}\left(\overline{Z}_{mt} - \overline{Z}_{mc}, \overline{W}_{mt} - \overline{W}_{mc}\right) = \boldsymbol{\psi}' \begin{bmatrix} \operatorname{var}\left(\overline{\mathbf{X}}^{(s)}_{mt} - \overline{\mathbf{X}}^{(s)}_{mc}\right) & \mathbf{C1}' \\ \mathbf{1C}' & k(\mathbf{I} + c_0\mathbf{1}\mathbf{1}') \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}^{(r)} \end{bmatrix}.$$

Using V_s to represent a vector with s components, the covariance can be written as

$$\begin{aligned}
&\operatorname{cov}\left(\overline{Z}_{mt} - \overline{Z}_{mc}, \overline{W}_{mt} - \overline{W}_{mc}\right) \\
&= \left[V_s', \boldsymbol{\psi}^{(s)\prime}\mathbf{C}\mathbf{1}' + k\boldsymbol{\psi}^{(r)\prime} + kc_0\boldsymbol{\psi}^{(r)\prime}\mathbf{1}\mathbf{1}'\right]\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\gamma}^{(r)}\end{bmatrix} \\
&= \boldsymbol{\psi}^{(s)\prime}\mathbf{C}\mathbf{1}'\boldsymbol{\gamma}^{(r)} + k\boldsymbol{\psi}^{(r)\prime}\boldsymbol{\gamma}^{(r)} + kc_0\boldsymbol{\psi}^{(r)\prime}\mathbf{1}\mathbf{1}'\boldsymbol{\gamma}^{(r)} \\
&= 0,
\end{aligned}$$

from (34) and (35) when $\boldsymbol{\mu}_t^{(r)} \neq \boldsymbol{\mu}_c^{(r)}$. When $\boldsymbol{\mu}_t^{(r)} = \boldsymbol{\mu}_c^{(r)}$, the result follows from (37) and (45), which implies $\mathbf{C} = \mathbf{0}$ and $c_0 = 0$. □

The proof of Corollary 4.4 is analogous to the proof of Corollary 3.4 with a modification like that in Corollary 4.3, and is not presented in detail.

5. Conclusions. Our theoretical framework has established three general results (Corollaries 3.2–3.4) concerning the performance of affinely invariant matching methods, and three general results (Corollaries 4.2–4.4) concerning the performance of conditionally affinely invariant matching methods. The first collection of results can be regarded as a special case of the latter results. They exhibit four ratios involving Z and three ratios involving W, all of which are free of dependence on the particular outcome variable Y, but do depend on the matching setting (i.e., distributional forms, sample sizes and the particular matching method employed). The latter collection of results exhibit four ratios involving $\mathscr{Z}$, which depend on Y as well as the matching setting, and three ratios involving $\mathscr{W}$, which are free of Y. Thus, the corollaries show that only a small number of quantities are needed to evaluate the sampling properties of complex matching procedures for a large class of theoretically important distributions.

Under normality, simple approximations for all ratios involving W and $\mathscr{W}$ can be obtained analytically for matching methods that use population or estimated discriminants. Furthermore, under these conditions, bounds can be obtained for the ratios involving Z and $\mathscr{Z}$. These results are presented in Rubin and Thomas (1991), and of particular importance, they describe the difference between using estimated and population discriminants. Subsequent work [Rubin and Thomas (1992)] using simulations and real data supports the relevance of these results to practice when distributions do not satisfy underlying assumptions.

Acknowledgments. We thank the referees for careful readings and thoughtful comments.

REFERENCES

CARPENTER, R. G. (1977). Matching when covariates are normally distributed. *Biometrika* **64** 299–307.

COCHRAN, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24** 295–313.

COCHRAN, W. G. and RUBIN, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā Ser. A* **35** 417–446.

DAUDIN, J. J. (1986). Selection of variables in mixed-variable discriminant analysis. *Biometrics* **42** 473–481.

DEMPSTER, A. P. (1969). *Continuous Multivariate Analysis*. Addison-Wesley, Reading, Mass.

EFRON, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.* **70** 892–898.

KRZANOWSKI, W. J. (1975). Discrimination and classification using both binary and continuous variables. *J. Amer. Statist. Assoc.* **70** 782–790.

KRZANOWSKI, W. J. (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics* **36** 486–499.

OLKIN, I. and TATE, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.* **32** 448–465.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55.

ROSSENBAUM, 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.

RUBIN, D. B. (1973a). Matching to remove bias in observational studies. *Biometrics* **29** 159–184.

RUBIN, D. B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* **29** 185–203.

RUBIN, D. B. (1976a). Multivariate matching methods that are equal percent bias reducing, I: Some examples. *Biometrics* **32** 109–120.

RUBIN, D. B. (1976b). Multivariate matching methods that are equal percent bias reducing, II: Maximums on bias reduction. *Biometrics* **32** 121–132.

RUBIN, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observation studies. *J. Amer. Statist. Assoc.* **74** 318–328.

RUBIN, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics* **36** 293–298.

RUBIN, D. B. and THOMAS, N. (1991). Characterizing the effect of matching using linear propensity score methods with normal covariates. *Biometrika*. To appear.

RUBIN, D. B. and THOMAS, N. (1992). Matched sampling using estimated linear propensity scores with ellipsoidal distributions. Technical report, Dept. Statistics, Harvard Univ.

DEPARTMENT OF STATISTICS
SCIENCE CENTER, HARVARD UNIVERSITY
ONE OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138

RESEARCH STATISTICS GROUP
EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541