Journal of Statistical Planning and Inference 140 (2010) 246 – 259

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Is regression adjustment supported by the Neyman model for causal inference?

Peter Z. Schochet*

Mathematica Policy Research, Inc., P.O. Box 2393, Princeton, NJ 08543-2393, USA

ARTICLE INFO

Article history:
Received 17 September 2007
Received in revised form
6 July 2009
Accepted 8 July 2009
Available online 15 July 2009

Keywords:
Neyman causal model
Experimental designs
Average treatment effects
Regression adjustment
Social policy interventions

ABSTRACT

This paper examines both theoretically and empirically whether the common practice of using OLS multivariate regression models to estimate average treatment effects (ATEs) under experimental designs is justified by the Neyman model for causal inference. Using data from eight large U.S. social policy experiments, the paper finds that estimated standard errors and significance levels for ATE estimators are similar under the OLS and Neyman models when baseline covariates are included in the models, even though theory suggests that this may not have been the case. This occurs primarily because treatment effects do not appear to vary substantially across study subjects.

1. Introduction

In the analysis of data from experimental designs, researchers typically estimate average treatment effects (ATEs) using multivariate regression models that control for baseline covariates. The use of multiple regression methods is typically advocated for two reasons. First, covariates can improve the statistical precision of the estimated ATEs to the extent that the covariates are correlated with the outcome measures. Second, covariates can adjust for differences between the treatment and control group's observable characteristics due to random selection and missing data. It is argued then, that if the functional form (behavioral) relationship between the outcomes and covariates is specified correctly, multiple regression methods produce unbiased estimates of ATEs, and yield standard errors that are smaller than simple differences-in-means ATE estimators.

There is a long tradition using ordinary least squares (OLS) regression models to analyze experimental data (see for example, Cochran and Cox, 1957; Fleiss, 1986; Byrk and Raudenbush, 1992). Freedman (2008) demonstrates, however, that this *reduced-form* regression modeling approach may not conform to the nonparametric model of causal inference for estimating ATEs introduced by Neyman (1923) and developed in Rubin (1974, 1978), Holland (1986), Heckman (2001, 2005), and Heckman and Vytlacil (2006). In particular, Freedman (2008) shows that the error structure assumed by the OLS model may not be supported by the *structural* regression model implied by randomization, and could yield biased asymptotic standard errors for ATE estimators.

These findings open up the question as to whether regression procedures are appropriate for analyzing experimental data. This paper aims to address this issue both theoretically and empirically using data from eight large, multi-site experimental evaluations of public programs and social policy interventions in the United States over the past 15 years. A contribution of this paper is the use of statistical procedures to estimate missing potential outcomes that are needed to apply the variance formulas

* Tel.: +1 609 936 2783; fax: +1 609 799 0005.
E-mail address: pschochet@mathematica-mpr.com (P.Z. Schochet).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2009.07.008

under the structural models. Another contribution is that it generalizes the asymptotic results in Freedman (2008) and Yang and Tsiatis (2001) to include multiple covariates (that are typically used to estimate regression-adjusted ATEs) rather than a single covariate. Finally, the paper considers both finite- and super-population models as well as stratified designs.

The focus is on randomized control trials (RCTs) with a single treatment and control group. It is assumed that the missing data process for study outcomes is independent of treatment status, so that the focus is on the use of regression models to increase precision rather than to adjust for nonresponse bias. The results in this paper pertain to OLS models (including linear probability models) but not to qualitative response models (such as logit or tobit models).

The rest of this paper is in five sections. Section 2 presents the Neyman nonparametric model for causal inference, and Section 3 discusses the super-population version of the model. Section 4 uses a unified regression approach to obtain ATE moment estimators under the structural and reduced-form models with and without covariates. Section 5 presents empirical results and Section 6 presents conclusions.

2. The Neyman finite-population model for causal inference

This paper considers an RCT where n subjects are randomly assigned to a treatment or control condition. The sample contains np treatments and $n(1-p)$ controls where p is the sampling rate to the treatment group ($0<p<1$). It is assumed for now that the n subjects define the population universe—the finite-population model considered by Neyman. Let Y_{Ti} be the "potential" outcome for subject i in the treatment condition and Y_{Ci} be the potential outcome for subject i in the control condition. The difference between the two fixed potential outcomes, $(Y_{Ti}-Y_{Ci})$, is the subject-level treatment effect, and the ATE parameter β_{ATE} (also known as the intention-to-treat [ITT] parameter) is the average treatment effect over all subjects:

$$\beta_{ATE}=\bar{Y}_T-\bar{Y}_C=\frac{1}{n}\sum_{i=1}^{n}(Y_{Ti}-Y_{Ci}). \tag{1}$$

The ATE parameter cannot be calculated directly because potential outcomes for each subject cannot be observed in both the treatment and control conditions. Formally, if T_i is a treatment status indicator variable that equals 1 for treatments and 0 for controls, then the *observed* outcome for a subject, y_i, can be expressed as follows:

$$y_i=T_iY_{Ti}+(1-T_i)Y_{Ci}. \tag{2}$$

In the Neyman model, potential outcomes are fixed and the only source of randomness is T_i. This model is nonparametric because it does not depend on the potential outcome distributions.

Following Freedman (2008), a *structural* regression model implied by the Neyman model can be constructed by re-writing (2) as follows:

$$y_i=\beta_0+\beta_{ATE}(T_i-p)+u_i, \tag{3}$$

where

$$\beta_0=p\bar{Y}_T+(1-p)\bar{Y}_C,\quad \beta_{ATE}=\bar{Y}_T-\bar{Y}_C, \tag{3a}$$

$$u_i=\alpha_i+\tau_i(T_i-p), \tag{3b}$$

$$\alpha_i=p(Y_{Ti}-\bar{Y}_T)+(1-p)(Y_{Ci}-\bar{Y}_C),\quad \tau_i=(Y_{Ti}-\bar{Y}_T)-(Y_{Ci}-\bar{Y}_C). \tag{3c}$$

The "error" term in this model, u_i, is a function of two terms: (1) α_i, the expected observed outcome for the subject; and (2) τ_i, the subject-level treatment effect. Note that over all n subjects, $\sum_i\alpha_i=\sum_i\tau_i=0$.

The structural model in (3) does not satisfy key assumptions of the OLS model—which is also referred to as the *reduced-form* model—because u_i does not have mean zero, and, to the extent that τ_i varies across subjects, u_i is heteroscedastic, (weakly) correlated across subjects, and correlated with the regressor (T_i-p):

$$E(u_i)=\alpha_i,\quad Var(u_i)=\tau_i^2p(1-p),$$

$$Cov(u_iu_j)=-\tau_i\tau_jp(1-p)/(n-1), \tag{4a}$$

$$E[(T_i-p_i)u_i]=\tau_ip(1-p). \tag{4b}$$

These potential breakdowns in the assumptions of the linear model account for the differences in asymptotic variance estimates using the Neyman and reduced-form models.

3. The super-population model

The super-population model assumes that the n subjects are a random sample from an infinite universe (Imbens and Rubin, 2001; Yang and Tsiatis, 2001). Under this model, study results are assumed to generalize beyond the study sample. Potential outcomes for the n subjects are assumed to be random draws from potential treatment and control outcome distributions in the super-population, with finite means and variances that are denoted by $\mu_{SP,T}$ and $\sigma^2_{SP,T}$ for potential treatment outcomes and $\mu_{SP,C}$ and $\sigma^2_{SP,C}$ for potential control outcomes. These two outcome distributions also define the distribution of subject-level treatment effects in the super-population, with mean $\mu_{SP,\tau}$ and variance $\sigma^2_{SP,\tau}$.

The ATE parameter under the super-population model is $\mu_{SP,\tau} = E(\bar{Y}_T - \bar{Y}_C) = \mu_T - \mu_C$. Thus, the structural regression model implied by this framework can be expressed as follows:

$$y_i = \mu_0 + \mu_{SP,\tau}(T_i - p) + \theta_i, \tag{5}$$

where

$$\mu_0 = p\mu_T + (1-p)\mu_C, \tag{5a}$$

$$\theta_i = \alpha_{i,SP} + \tau_{i,SP}(T_i - p), \tag{5b}$$

$$\alpha_{i,SP} = p(Y_{Ti} - \mu_T) + (1-p)(Y_{Ci} - \mu_C), \quad \tau_{i,SP} = (Y_{Ti} - \mu_T) - (Y_{Ci} - \mu_C). \tag{5c}$$

Under this model, T_i, Y_{Ti}, Y_{Ci} and τ_i are random variables. Because of random assignment, $E(T_iY_{Ti}) = E(T_iY_{Ci}) = E(T_i\tau_i) = 0$. In addition, because $E(\alpha_{i,SP}) = E(\tau_{i,SP}) = 0$, the error term θ_i has mean zero and is uncorrelated with $(T_i - p)$:

$$E(\theta_i) = E(\alpha_{i,SP}) + E[\tau_{i,SP}(T_i - p)|T_i = 1)p + E[\tau_{i,SP}(T_i - p)|T_i = 0)(1-p) = 0, \tag{6a}$$

$$E[(T_i - p)\theta_i] = E[(T_i - p)\theta_i|T_i = 1]p + E[(T_i - p)u_i|T_i = 0](1-p) = 0 \tag{6b}$$

Furthermore, θ_i has constant variance across subjects within research condition and is uncorrelated across subjects:

$$Var(\theta_i|T_i = 1) = E[[\alpha_{i,SP} + \tau_{i,SP}(T_i - p)]^2|T_i = 1] = \sigma^2_{T,SP} \quad \text{and} \quad Var(\theta_i|T_i = 0) = \sigma^2_{C,SP}, \tag{6c}$$

$$Cov(\theta_i\theta_{i'}) = E(\theta_i\theta_{i'}) = 0. \tag{6d}$$

Thus, the super-population model satisfies the usual OLS assumptions except that error variances are heteroscedastic across the two research groups.

4. Moments under the structural and reduced-form models

The models that this paper considers are versions of the following generic regression model:

$$\mathbf{y} = \gamma_1\tilde{\mathbf{T}} + \tilde{\mathbf{Z}}\boldsymbol{\gamma}_2 + \mathbf{e}, \tag{7}$$

where $\mathbf{y}$ is an $n\times1$ vector of observed outcomes; $\tilde{\mathbf{T}}$ is an $n\times1$ vector whose ith element is $(T_i - p)$; $\tilde{\mathbf{Z}} = [\mathbf{K}\ \ \tilde{\mathbf{X}}]$ is an $n\times k$ real-valued matrix with full column rank, where $\mathbf{K}$ is a column of 1 s for the intercept and $\tilde{\mathbf{X}}$ is a matrix of *centered* baseline covariates ($\tilde{X}_{ij} = X_{ij} - \bar{X}_j$ for subject i and covariate j); $\gamma_1(1\times1)$ and $\boldsymbol{\gamma}_2(k\times1)$ are parameters to be estimated; and $\mathbf{e}$ ($n\times1$) is a generic error term. The least squares estimator for γ_1 is as follows:

$$\hat{\gamma}_1 = [\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{Z}}})\tilde{\mathbf{T}}]^{-1}\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{Z}}})\mathbf{y} = [\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}})\tilde{\mathbf{T}}]^{-1}\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}})\mathbf{y}, \tag{8}$$

where $\mathbf{I}$ is the $n\times n$ identity matrix and $\mathbf{P}_{\tilde{\mathbf{Z}}}$ and $\mathbf{P}_{\tilde{\mathbf{X}}}$ are projection matrices of the form $\mathbf{P}_\mathbf{M} = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$. The second equality in (8) holds because $\mathbf{P}_{\tilde{\mathbf{Z}}}=\mathbf{P}_\mathbf{K}+\mathbf{P}_{\tilde{\mathbf{X}}}$ and $\tilde{\mathbf{T}}'\mathbf{P}_\mathbf{K} = 0$.

4.1. Models without covariates

4.1.1. Structural finite- and super-population models

This section presents two well-known lemmas for the finite- and super-population models. Proofs are provided in Appendix A using the regression frameworks in (3) and (5). Alternative proofs are provided in Freedman (2008) for Lemma 1, Imbens and Rubin (2001) for Lemmas 1 and 2, and Yang and Tsiatis (2001) for Lemma 2. Proofs are provided because they form the basis for the proofs pertaining to the models with covariates.

Lemma 1. *Let $b_{SR,FP}$ be the simple regression estimator for β_{ATE} under the finite-population model in* (3). *Then, $b_{SR,FP} = (\bar{y}_T - \bar{y}_C)$ is unbiased with variance*:

$$Var(b_{SR,FP}) = \frac{S_T^2}{np} + \frac{S_C^2}{n(1-p)} - \frac{S_\tau^2}{n}, \tag{9}$$

where

$$S_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{Ti}-\bar{Y}_T)^2, \quad S_C^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{Ci}-\bar{Y}_C)^2, \quad S_\tau^2 = \frac{1}{n-1}\sum_{i=1}^{n}\tau_i^2.$$

Furthermore, as n increases to infinity for an increasing sequence of finite populations, assume that:

$$S_T^2 \to \bar{S}_T^2, \quad S_C^2 \to \bar{S}_C^2, \quad S_\tau^2 \to \bar{S}_\tau^2, \tag{10}$$

where $\bar{S}_T^2, \bar{S}_C^2$, *and* $\bar{S}_\tau^2$ *are fixed, nonnegative, real numbers. Then*, $b_{\mathrm{SR,FP}}$ *is asymptotically normal with variance*:

$$AsyVar(b_{SR,FP}) = \frac{\bar{S}_T^2}{np} + \frac{\bar{S}_C^2}{n(1-p)} - \frac{\bar{S}_\tau^2}{n}. \tag{11}$$

Lemma 2. *Let* $b_{SR,SP}$ *be the simple regression estimator for* $\mu_{SR,SP}$ under *the super-population model in* (5). *Then*, $b_{SR,SP} = (\bar{y}_T - \bar{y}_C)$ *is unbiased and asymptotically normal with variance*:

$$Var(b_{SR,SP}) = \frac{\sigma_{SP,T}^2}{np} + \frac{\sigma_{SP,C}^2}{n(1-p)}. \tag{12}$$

Discussion. Unbiased estimates for S_T^2 and S_C^2 in (9), $\bar{S}_T^2$ and $\bar{S}_C^2$ in (11), and $\sigma_{SP,T}^2$ and $\sigma_{SP,C}^2$ in (12) can be obtained using sample variances for the treatment and control groups, s_T^2 and s_C^2, respectively:

$$s_T^2 = \frac{1}{np-1}\sum_{i:T_i=1}^{np}(y_i-\bar{y}_T)^2, \quad s_C^2 = \frac{1}{n(1-p)-1}\sum_{i:T_i=0}^{n(1-p)}(y_i-\bar{y}_C)^2.$$

Estimators for S_τ^2 in (9) and $\bar{S}_\tau^2$ in (11) take the form $s_\tau^2 = \sum_{i=1}^{n}\hat{\tau}_i^2/(n-1)$, where $\hat{\tau}_i$ is an estimate of the treatment effect for the ith subject (see below). The variance estimator is larger under the super-population model than the finite-population model (by s_τ^2/n), because the estimated ATEs are externally valid under the super-population model but internally valid only under the finite-population model.

4.1.2. Reduced-form model and comparisons to the structural models

Under the reduced-form model, the regression model in (7) is estimated under the usual OLS assumptions, where e_i is assumed to be a generic mean zero, homoscedastic, random error term with variance $\sigma_{SR,RF}^2$, that is pairwise uncorrelated across subjects and uncorrelated with treatment status. Although not usually explicitly stated, the relevant ATE parameter under the reduced-form model is the super-population ATE parameter.

The simple regression estimator under the reduced-form model—denoted by $b_{SR,RF}$—is $\bar{y}_T - \bar{y}_C$. Standard OLS methods can be used to show that $b_{SR,RF}$ is unbiased for μ_{ATE}, has variance equal to $\sigma_{SR,RF}^2/np(1-p)$, and is asymptotically normal. The variance estimator for $b_{SR,RF}$ is as follows:

$$\hat{Var}(b_{SR,RF}) = \frac{s_{Pooled}^2}{np(1-p)}, \quad s_{Pooled}^2 = \frac{1}{n-2}[s_T^2(np-1) + s_C^2(n(1-p)-1)],$$

where s_{Pooled}^2 is an unbiased estimator for $\sigma_{SR,RF}^2$ based on the residual sum of squares obtained using the pooled treatment and control groups.

This reduced-form variance estimator is justified under the super-population model if $\sigma_{SP,T}^2 = \sigma_{SP,C}^2 = \sigma_{SP}^2$. In this case, σ_{SP}^2 can be estimated using s_{Pooled}^2. Even if the two variances are not the same, however, the reduced-form and super-population estimators will be identical if $p = 0.5$, which is a common design. The reduced-form model is also justified under both the finite- and super-population models if subject-level treatment effects are constant. In this case, $S_\tau^2 = 0$ and $S_T^2 = S_C^2 = S_{Pooled}^2$ which can be estimated by s_{Pooled}^2. With heterogeneous treatment effects, the variance estimator under the reduced-form model will be larger than under the finite-population model if $p = 0.5$, but could be larger or smaller than under the finite-population model if $p \neq 0.5$.

4.2. Models with covariates

This section examines asymptotic moments for ATE estimators when the models include covariates $\tilde{\mathbf{X}}$ pertaining to the pre-randomization period. Because of randomization, $\tilde{\mathbf{X}}$ is not indexed by T or C. The covariates could include pre-intervention measures of the outcomes and could be binary or continuous. It is assumed initially that subjects are randomly assigned independently of $\tilde{\mathbf{X}}$, but stratified designs are considered later in this paper.

In the structural models, the covariates are considered to be irrelevant variables because (3) and (5) are the true models. Thus, the ATE parameters considered above for the structural models without covariates pertain *also* to the models with covariates. In the reduced form model, however, the true behavioral model is assumed to include the covariates.

4.2.1. Structural finite-population model

To examine asymptotic moments under the finite-population model with fixed covariates, it is assumed in addition to (10) that as n approaches infinity:

$$\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n} \to \mathbf{\Omega_{XX}}, \quad \frac{\tilde{\mathbf{X}}'\boldsymbol{\alpha}}{n} \to \mathbf{\Omega_{X\alpha}}, \quad \frac{\tilde{\mathbf{X}}'\boldsymbol{\tau}}{n} \to \mathbf{\Omega_{X\tau}}, \tag{13}$$

where $\boldsymbol{\alpha}$ is an $n\times 1$ column vector containing the $\alpha_i s$; $\boldsymbol{\tau}$ is an $n\times 1$ column vector containing the $\tau_i s$; $\mathbf{\Omega_{XX}}$ is an $n\times n$ symmetric, finite, positive definite matrix; and $\mathbf{\Omega_{X\alpha}}$ and $\mathbf{\Omega_{X\tau}}$ are finite $k\times 1$ vectors of fixed real numbers. In these expressions, the covariances between the covariates and potential outcomes can differ for treatments and controls.

The following lemma generalizes results in Freedman (2008) to allow for multiple covariates rather than a single covariate. The proof is provided in Appendix A.

Lemma 3. *Let $b_{MR,FP}$ be the multiple regression estimator for β_{ATE} under the structural model in* (3) *and assume* (10) *and* (13). *Then, $b_{MR,FP}$ is asymptotically normal with asymptotic mean β_{ATE} and asymptotic variance*:

$$AsyVar(b_{MR,FP}) = \left(\frac{\bar{S}_T^2}{np} + \frac{\bar{S}_C^2}{n(1-p)} - \frac{\bar{S}_\tau^2}{n}\right) - \frac{\mathbf{\Omega}'_{\mathbf{X}\alpha}\mathbf{\Omega}^{-1}_{\mathbf{XX}}\mathbf{\Omega}_{\mathbf{X}\alpha}}{np(1-p)} - 2(1-2p)\frac{\mathbf{\Omega}'_{\mathbf{X}\tau}\mathbf{\Omega}^{-1}_{\mathbf{XX}}\mathbf{\Omega}_{\mathbf{X}\alpha}}{np(1-p)}. \tag{14}$$

Discussion. The asymptotic variance in (14) can be estimated as follows:

$$Asy\hat{V}ar(b_{MR,FP}) = \left(\frac{s_T^2}{np} + \frac{s_C^2}{n(1-p)} - \frac{s_\tau^2}{n}\right) - \frac{(1/n)\hat{\boldsymbol{\alpha}}'\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\alpha}}}{np(1-p)} - 2(1-2p)\frac{(1/n)\hat{\boldsymbol{\tau}}'\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\alpha}}}{np(1-p)}, \tag{15a}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\tau}}$ are estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$, respectively. The first bracketed term on the right-hand side in (15a) is the variance estimator under the finite-population model without covariates, so the second and third terms represent precision gains or losses from adding covariates. The numerator in the second term is the mean explained sum of squares from a regression of $\hat{\boldsymbol{\alpha}}$ on $\tilde{\mathbf{X}}$. This term is nonnegative, and hence, can only generate precision gains. The numerator in the third term is $-2(1-2p)$ times the covariance of predicted values from regressions of $\hat{\boldsymbol{\tau}}$ on $\tilde{\mathbf{X}}$ (that is, $\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\tau}}$) and of $\hat{\boldsymbol{\alpha}}$ on $\tilde{\mathbf{X}}$ (that is, $\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\alpha}}$). This term could be negative, positive, or zero.

An alternative estimation approach is to estimate the covariances in (14) using sample moments:

$$\hat{\mathbf{\Omega}}_{\mathbf{X}\alpha} = p\mathbf{H_T} + (1-p)\mathbf{H_C}, \quad \hat{\mathbf{\Omega}}_{\mathbf{X}\tau} = \mathbf{H_T} - \mathbf{H_C}, \quad \hat{\mathbf{\Omega}}_{\mathbf{XX}} = \tilde{\mathbf{X}}'\tilde{\mathbf{X}}/n, \tag{15b}$$

where $\mathbf{H_T}$ and $\mathbf{H_C}$ are $k\times 1$ vectors of sample covariances between $\tilde{\mathbf{X}}$ and $\mathbf{y}$ for treatments and controls, respectively:

$$\mathbf{H_T}(j) = \frac{1}{np-1}\sum_{i:T_i=1}^{np}(X_{ij}-\bar{X}_{Tj})(y_i-\bar{y}_T),$$

$$\mathbf{H_C}(j) = \frac{1}{n(1-p)-1}\sum_{i:T_i=0}^{n(1-p)}(X_{ij}-\bar{X}_{Cj})(y_i-\bar{y}_C).$$

For many scenarios, regression adjustment will produce precision gains under the structural model. First, precision gains will occur if $p=0.5$ because the last covariance term in (14) vanishes in this case. Second, for the same reason, precision gains will occur under constant treatment effects. Third, even with heterogeneous treatment effects, precision gains will occur if the correlations are low between τ_i and $\tilde{X}_{ij}$. Finally, precision gains will occur if $p<0.5$ and the covariance term is positive or if $p>0.5$ and the covariance term is negative. The only way for precision losses to occur is if the third term is negative and larger than the second term.

4.2.2. Structural super-population model

Under the super-population model with covariates, the covariates as well as the potential outcomes are considered to be random draws from joint super-population distributions. For this model, the covariates tend to be correlated with the error term. The proof of the following lemma is in Appendix A.

Lemma 4. *Let $b_{MR,SP}$ be the multiple regression estimator for μ_{ATE} under the structural model in* (5). *Then, $b_{MR,SP}$ is asymptotically normal with mean μ_{ATE} and variance*:

$$AsyVar(b_{MR,SP}) = \left(\frac{\sigma_{SP,T}^2}{np} + \frac{\sigma_{SP,C}^2}{n(1-p)}\right) - \frac{\mathbf{\Lambda}'_{\mathbf{X}\alpha}\mathbf{\Lambda}^{-1}_{\mathbf{XX}}\mathbf{\Lambda}_{\mathbf{X}\alpha}}{np(1-p)} - 2(1-2p)\frac{\mathbf{\Lambda}'_{\mathbf{X}\tau}\mathbf{\Lambda}^{-1}_{\mathbf{XX}}\mathbf{\Lambda}_{\mathbf{X}\alpha}}{np(1-p)} \tag{16}$$

where $\mathbf{\Lambda_{XX}}$, $\mathbf{\Lambda_{X\alpha}}$, and $\mathbf{\Lambda_{X\tau}}$ are respective moment matrices for $\mathbf{X}^{*\prime}\mathbf{X}^*$, $\mathbf{X}^{*\prime}\boldsymbol{\alpha}$, and $\mathbf{X}^{*\prime}\boldsymbol{\tau}$ under the joint super-population distribution for the covariates and potential outcomes, where $X^*_{ij} = X_{ij} - E(X_{ij})$.

Discussion. The variance in (16) can be estimated as follows:

$$Asy\hat{V}ar(b_{MR,SP}) = \left(\frac{s_T^2}{np} + \frac{s_C^2}{n(1-p)}\right) - \frac{(1/n)\hat{\boldsymbol{\alpha}}'\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\alpha}}}{np(1-p)} - 2(1-2p)\frac{(1/n)\hat{\boldsymbol{\tau}}'\mathbf{P}_{\tilde{\mathbf{X}}}\hat{\boldsymbol{\alpha}}}{np(1-p)}. \quad (17)$$

Alternatively, $\boldsymbol{\Lambda}_{\mathbf{XX}}$, $\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}}$, and $\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\tau}}$ can be estimated using sample moments (see (15b)). These estimators are larger than the corresponding ones for the finite-population model by s_τ^2/n.

4.2.3. Structural stratified designs

Under stratified experimental designs, the Neyman framework is applied within exogenous strata. A common approach for estimating ATEs under these designs is to include $\tilde{\mathbf{X}}$-by-$\tilde{\mathbf{T}}$ interaction terms as explanatory variables in the regression models in addition to $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{T}}$, where $\tilde{\mathbf{X}}$ contains binary strata indicators (but no other covariates). Let $\tilde{\mathbf{Q}}$ signify these interaction terms where $\tilde{Q}_{ij} = \tilde{X}_{ij}\tilde{T}_i$ for subject i in strata j. The asymptotic properties of estimators for this model are considered next. For ease of presentation, the focus is on the super-population model; differences between results for the finite- and super-population models are similar to those from above. The following lemma is proved in Appendix A.

Lemma 5. *Let $b_{MR,SP,STR}$ be the multiple regression estimator for μ_{ATE} under a stratified design using the super-population model in* (5) *with explanatory variables $\tilde{\mathbf{Z}} = [\mathbf{K}\ \ \tilde{\mathbf{X}}\ \ \tilde{\mathbf{Q}}]$. Then, $b_{MR,SP,STR}$ is asymptotically normal with mean μ_{ATE} and variance*:

$$AsyVar(b_{MR,SP,STR}) = \left(\frac{\sigma_{SP,T}^2}{np} + \frac{\sigma_{SP,C}^2}{n(1-p)}\right) - \frac{\boldsymbol{\Lambda}'_{\mathbf{X}\boldsymbol{\alpha}^*}\boldsymbol{\Lambda}_{\mathbf{XX}}^{-1}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}^*}}{np(1-p)}, \quad (18)$$

where $\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}^*}$ is the moment matrix for $\mathbf{X}^{*\prime}\boldsymbol{\alpha}^*$ and $\alpha_i^* = (1-p)(Y_{Ti} - \mu_T) + p(Y_{Ci} - \mu_C)$.

Discussion. Because $\boldsymbol{\Lambda}'_{\mathbf{X}\boldsymbol{\alpha}^*}\boldsymbol{\Lambda}_{\mathbf{XX}}^{-1}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}^*}$ is nonnegative, regression adjustment under the stratified version of the Neyman model can only yield precision gains. Furthermore, because $\alpha^* - \alpha = (1-2p)\tau$, it can be shown that the variance in (18) is less than or equal to the variance in (16). Note that (18) holds more generally (under both stratified and non-stratified designs) if the models include interactions between $\tilde{\mathbf{T}}$ and *non-strata-related* covariates. However, models with these interactions are not structural models implied by the Neyman model, but are reduced-form models where treatment effects are assumed *ex post* to differ across covariate levels.

4.2.4. Reduced-form model and comparisons to the structural models

Under the reduced-form model with covariates, the regression model in (7) is estimated under the usual OLS assumptions, where the covariates are assumed to be uncorrelated with the error term. Let $\tilde{\mathbf{Z}}^{OLS} = [\mathbf{K}\,\tilde{\mathbf{T}}\,\tilde{\mathbf{X}}]$ be the model explanatory variables, and let $b_{\text{MR,RF}}$ be the multiple regression estimator under the reduced-form model in (7). Then $b_{\text{MR,RF}}$ is unbiased with the following variance estimator:

$$\hat{Var}(b_{MR,RF}) = \left[\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{Z}}^{OLS}})\tilde{\mathbf{T}}\right]^{-1} s_{MR,RF}^2; \quad s_{MR,RF}^2 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{Z}}^{OLS}})\mathbf{y}/(n-k-1).$$

Because $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{T}}$ are asymptotically uncorrelated, $\mathbf{P}_{\tilde{\mathbf{Z}}^{OLS}}$ can be approximated by $\mathbf{P}_{\mathbf{K}} + \mathbf{P}_{\tilde{\mathbf{T}}} + \mathbf{P}_{\tilde{\mathbf{X}}}$. Thus, an asymptotic approximation to the OLS variance estimator is

$$Asy\hat{V}ar(b_{MR,RF}) = \frac{s_{Pooled}^2}{np(1-p)} - \frac{(1/n)\mathbf{y}'\mathbf{P}_{\tilde{\mathbf{X}}}\mathbf{y}}{np(1-p)}. \quad (19)$$

The variance in (19) is used for the empirical analysis to consistently compare variance estimators under the reduced-form and structural models.

Discussion. A key issue for this article is how the finite- and super-population variance estimators in (15a) and (17) compare to the reduced-form variance estimator in (19). The reduced-form and super-population estimators will be alike if the variances of y_i and the covariances between y_i and $\tilde{X}_{ij}$ are similar for treatments and controls (that is, if the covariances between τ_i and $\tilde{X}_{ij}$ are small). Even if the two sets of variances and covariances differ, however, the reduced-form model is largely justified by the super-population model if $p = 0.5$. Furthermore, under constant treatment effects, the three models are identical because in this case: (1) $S_\tau^2 = 0$; (2) $S_T^2 = S_C^2 = S_{Pooled}^2$; (3) $\boldsymbol{\alpha}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} = \mathbf{y}'\mathbf{P}_{\tilde{\mathbf{X}}}\mathbf{y}$; and (4) $\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} = 0$. For other cases, the variance estimators for the structural models could be smaller or larger than for the reduced-form model.

5. Empirical analysis

This section uses data from eight large experimental social policy evaluations to address the following two research questions: (1) in practice, are there precision gains from using covariates under the Neyman structural model? and (2) in practice, do variance estimators for ATEs differ under the reduced-form and structural models? The stratified model is not examined because some of the considered studies did not use stratified designs.

5.1. Analytic methods

Several approaches were used to calculate the variance estimators under the structural models. First, two methods were used to impute missing potential outcomes to obtain estimates of α_i and τ_i, and hence, estimates of s_τ^2, $\boldsymbol{\alpha}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}$, and $\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}$:

(1) *Subgroup method*. ATEs were estimated for a large number of subgroups by regressing y_i on terms formed by interacting $\tilde{T}_i$ with baseline covariates. Estimates for τ_i were obtained using predicted ATEs from these subgroup interaction models. This approach cannot provide estimates for $\boldsymbol{\alpha}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}$ or $\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}$ because, by construction, estimates of α_i and τ_i are perfectly correlated with the covariates. However, this approach was used to construct estimates for s_τ^2 that are based on actual subgroup ATEs observed under the experimental design. This approach may underestimate s_τ^2 because the estimates for τ_i are constant within subgroup cells, and the covariates may not fully explain the variation in τ_i. Smith et al. (2005) used a similar approach in a different context.

(2) *Rank-order method*. This method assumes the same rank order for treatment and control potential outcomes. For example, if the observed outcome for a treatment subject is at the 75th percentile of the outcome distribution for the treatment group, then it is assumed that the potential control outcome for that subject is at the 75th percentile of the observed outcome distribution for the control group. This procedure identifies potential outcomes without relying on baseline covariates, and may represent a reasonable approximation for interventions like the ones that are considered where ATEs are small relative to the mean outcomes. This approach draws on the literature on the estimation of quantile treatment effects (Abadie et al., 2002). Heckman et al. (1997) used this approach in a different context. The imputations were performed within strata to improve their quality (see Schochet, 2007).

The final *Covariance method* involved using (15b) to directly estimate the covariance terms in (14) and (16) without having to first impute missing potential outcomes.

The variance estimators under the reduced-form and structural models are identical if treatment effects are constant across subjects. Thus, to further assess the appropriateness of the reduced-form model, the analysis examined parameter restrictions in the variance formulas that are implied by homogeneous treatment effects: (1) $s_T^2 = s_C^2 = s_{Pooled}^2 = s_\alpha^2$; (2) $s_\tau^2 = 0$, which implies that $\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} = 0$; and (3) the R^2 value when $\mathbf{y}$ is regressed on $\tilde{\mathbf{X}}$ ($R^2_{y;X}$) is the same for treatments and controls. In addition, the reduced-form model variance estimates were compared to White's heteroscedasticity-corrected variance estimates (White, 1980).

Most datasets used for the analysis contain similar numbers of treatment and control subjects. This has implications for the analysis because when $p = 0.5$ precision gains from regression adjustment will occur under the Neyman model. Because the situation is somewhat more complex when $p \neq 0.5$, the variance formulas were also estimated using $p = 1/3$ and $p = 2/3$ (which are realistic bounds used in RCTs) by randomly subsampling study subjects.

5.2. Data

Data for the analysis come from eight large experimental evaluations of U.S. public programs and social policy interventions conducted by Mathematica Policy Research, Inc. over the past 15 years. The studies were funded either by federal agencies or foundations and most had national advisory panels of evaluation and subject-area experts. These studies were selected because of their policy importance, and because they cover a wide range of policy areas and study populations in education, early childhood, labor, welfare, and health. Subjects in each study were randomly assigned to a single treatment or control group and the studies collected a rich set of outcome measures and baseline covariates that were used for regression adjustment. The study random assignment procedures were well-implemented and data collection response rates were reasonably high. Schochet (2007) provides a detailed description of the evaluations, including data sources, key outcomes, baseline covariates, and sampling strata. A brief description of the RCTs is provided below in the presentation of the impact findings.

5.3. Results

5.3.1. Summary of ate findings across the studies

Table 1 presents key impact findings for sixteen outcome measures across the eight studies based on estimates from the reduced-form approach without covariates and with covariates (the primary estimates used by the studies). The table denotes statistical significance at both the 5 and 10 percent significance levels because most studies used this standard. For the models with covariates, estimated impacts are statistically significant for ten of sixteen outcomes. A brief summary of findings is as follows:

- Providing vouchers to private schools of up to $1,400 per year did not change the test scores of low-income New York City students even though about three-quarters of the treatment students used the vouchers (Mayer et al., 2002).
- Average third-grade math test scores in low-income schools were significantly higher for treatment students taught by Teach for America teachers (recent graduates from selective colleges and universities) than control students taught by

Table 1
Key study results based on the reduced-form model.

Evaluation and outcome measure	Control group mean	No covariates in model		Covariates in model		
		Estimated ATE	Standard error	Estimated ATE	Standard error	Regression R^2 value
NYC Vouchers						
Reading scores (percentile points)	25.0	0.8	1.223	1.3	0.962	0.40
Math scores (percentile points)	25.1	−1.6	1.357	−1.3	1.133	0.27
Teach for America						
Math scores (percentile points)	30.3	0.8	0.910	2.0*	0.653	0.48
Reading scores (percentile points)	26.7	−0.4	1.017	0.2	0.578	0.66
21st Century						
Reading scores (percentile points)	41.3	−0.5	1.246	−0.3	1.159	0.30
Math scores (percentile points)	52.2	0.4	1.721	0.5	1.677	0.25
Disciplinary problems index	1.7	0.1	0.039	0.1	0.038	0.12
Job Corps						
Year 4 earnings per week ($s)	197.1	11.8*	3.823	15.3*	3.624	0.12
Number of arrests	0.76	−0.14*	0.028	−0.08*	0.026	0.14
Early Head Start						
Bayley MDI scores	90.0	1.1**	0.621	1.5*	0.552	0.25
HOME scores	27.0	0.5*	0.225	0.5*	0.193	0.29
San Diego Cash-Out						
Value of purchased foods ($s per week)	71.4	−6.4*	2.570	−5.8*	2.263	0.27
Food energy as a percentage of the RDA	140.2	−7.4*	3.774	−6.8**	3.733	0.04
Teenage Parent Demonstration						
% of months in work/school activities	27.8	7.4*	1.209	7.2*	1.158	0.10
Monthly earnings ($)	26.6	5.1*	2.172	4.5*	2.083	0.11
Cash and Counseling						
All Medicaid expenditures ($)	326.8	32.3*	10.111	29.3*	6.952	0.58

Note: see Appendix Table B.1 in Schochet (2007) for study details.
*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.10 level, two-tailed test.

traditionally-trained teachers (many of whom had many years of teaching experience), although there were no differences in reading scores (Decker et al., 2004).

- There were no statistically significant differences in test scores and disciplinary behavior measures between elementary school students who were offered the opportunity to enroll in after-school programs (some of which had an academic focus) and those who were not (Dynarski et al., 2004).
- Participation in Job Corps (the largest federal education and training program for disadvantaged youths between the ages of 16 and 24) increased earnings by about $1,000 in the fourth year after random assignment and reduced arrest rates (Schochet et al., 2008).
- Participation in the Early Head Start program—which provides high-quality child development and family services to children ages 0–3 at home or in child care centers—increased cognitive child development scores and the quality of the home environment (Love et al., 2002).
- Providing checks in place of food stamp program (FSP) coupons to FSP participants in San Diego significantly decreased (1) the money value of food purchased at home and (2) food energy consumed as a percentage of the recommended dietary allowance (RDA) (Ohls et al., 1992).
- Teenage first-time welfare mothers in the treatment group who were required to participate in employment, job training, and education activities in order to receive welfare (and who were also provided child care and transportation assistance) spent considerably more time in activities and earned more than control group members during the first two years after random assignment (although these effects disappeared after the demonstration ended) (Maynard et al., 1993).
- Among Medicaid beneficiaries eligible for personal care services, treatment group members who were provided a monthly allowance (paid by Medicaid) to hire their own caregivers had higher Medicaid costs one year after random assignment than control group members who received traditional services through home care agencies. This occurred because treatments were much more satisfied than controls with the services they received, and hence, used more services (Dale and Brown, 2005).

There are several notable features about the study results presented in Table 1 that inform the analysis presented below. First, the ATE estimators—that pertain to both the reduced-form and structural models—are similar based on models that include and exclude baseline covariates. Under the structural models, regression-adjusted ATE estimators are biased in small samples, but the biases appear to be small in the considered studies due to large sample sizes. Second, even for the ATE estimates that

Table 2
ATE standard errors and statistical significance levels for structural and reduced-form models.

Evaluation and outcome measure	No covariates in model			Covariates in model			
	Finite-population	Super-population	Reduced-form	Finite-population		Super-population	
				Covariance method	Rank-order method	Rank-order method	Reduced-form: asymptotic
NYC Vouchers							
Reading scores (percentile points)	1.263	1.273	1.223	0.923	0.971	1.060	0.978
Math scores (percentile points)	1.409	1.422	1.357	1.180	1.161	1.204	1.153
Teach for America							
Math scores (percentile points)	0.904	0.905	0.910	0.649*	0.642*	0.692*	0.643*
Reading scores (percentile points)	1.023	1.025	1.017	0.622	0.554	0.655	0.569
21st Century							
Reading scores (percentile points)	1.234	1.260	1.246	0.981	1.037	1.148	1.097
Math scores (percentile points)	1.682	1.754	1.721	1.421	1.512	1.645	1.576
Disciplinary problems index	0.039	0.039	0.039	0.036	0.037	0.038	0.038
Job Corps							
Year 4 earnings per week ($s)	3.771*	3.780*	3.823*	3.475*	3.518*	3.600*	3.603*
Number of arrests	0.030*	0.030*	0.028*	0.026*	0.027*	0.028*	0.027*
Early Head Start							
Bayley MDI scores	0.611**	0.623**	0.621**	0.522*	0.524*	0.533*	0.542*
HOME scores	0.233*	0.237*	0.225*	0.184*	0.193*	0.201*	0.199*
San Diego Cash-Out							
Value of purchased foods ($s per week)	2.560*	2.571*	2.570*	2.176*	2.219*	2.306*	2.238*
Food energy as a percentage of the RDA	3.765*	3.777*	3.774*	3.645**	3.631**	3.697**	3.692**
Teenage Parent Demonstration							
% of months in work/school activities	1.204*	1.209*	1.209*	1.138*	1.144*	1.158*	1.150*
Monthly earnings ($)	2.196*	2.201*	2.172*	2.024*	2.088*	2.098*	2.096*
Cash and Counseling							
All Medicaid expenditures ($)	10.083*	10.112*	10.111*	4.552*	6.919*	8.873*	6.926*

Note: see Appendix Table B.1 in Schochet (2007) for study details and the text for variance formulas. Estimates are based on actual study sample sizes. The s_τ^2 values for the finite-population models were obtained using the subgroup imputation method for models without covariates, and the rank-order imputation method for models with covariates.
*ATE estimate is significantly different from zero at the 0.05 level, two-tailed test.
**ATE estimate is significantly different from zero at the 0.10 level, two-tailed test.

are statistically significant, the estimated impacts are relatively small. This often occurs in social policy experiments where interventions are compared to the status quo condition, and hence, can realistically make marginal improvements only. Third, results presented in the study reports indicate that the estimated ATEs do not vary much across subgroups (not shown). Fourth, there is a wide range of R^2 values across the studies (from 0.02 to 0.66), so that the potential gains from regression vary by study; the R^2 values tend to be higher in models where the covariates contain pre-intervention measures of the outcome. Finally, statistical significance levels are similar in models that include and exclude covariates (except for the Teach for America study that has relatively small samples and high R^2 values).

5.3.2. Is regression adjustment supported by the data?

The data provide evidence that the OLS approach is justified by the Neyman model. The estimated standard errors and statistical significance levels for the estimated ATEs are similar for the finite-population, super-population, and reduced-form models with and without covariates (Table 2). Precision gains from regression adjustment are observed under the structural models and are similar to those observed under the reduced-form models (Tables 2 and 3). In addition, precision gains are similar for various values of p (Table 3).

These results hold for several reasons (Tables 4 and 5). First, treatment effects do not appear to vary substantially across subjects. Based on both the subgroup and rank-order methods, s_τ^2 values are modest relative to s_{Pooled}^2 values (Table 4). The s_τ^2 values become even smaller in the variance expressions because they are deflated by n whereas s_{Pooled}^2 values are deflated by $np(1-p)$. Furthermore, $R_{y;X}^2$ values are typically similar for treatments and controls, White's heteroescedacity-corrected standard errors are similar to the reduced-form standard errors, the values of $[(1/n)\boldsymbol{\alpha}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}/s_{Pooled}^2]$ are similar to $R_{y;X}^2$ values, and although s_T^2 values tend to differ somewhat from s_C^2 values, the differences are not large (Table 5). A second, related reason for the results is that values for $[(1/n)\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}/s_{Pooled}^2]$ and $[(1/n)\boldsymbol{\tau}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\tau}/s_{Pooled}^2]$ are small, suggesting low correlations between the subject-level treatment effects and the covariates (Table 4).

Table 3
Percentage reductions in ATE variance estimates from including covariates in the regression models.

Evaluation and outcome measure	Actual value for p		Alternative values for p for finite-population model	
	Finite-population	Reduced-form asymptotic	$p=1/3$	$p=2/3$[a]
NYC Vouchers				
Reading scores (percentile points)	40	41	45	39
Math scores (percentile points)	32	33	34	30
Teach for America				
Math scores (percentile points)	50	50	49	49
Reading scores (percentile points)	71	69	72	68
21st Century				
Reading scores (percentile points)	29	23	31	22
Math scores (percentile points)	19	17	21	16
Disciplinary problems index	8	8	8	8
Job Corps				
Year 4 earnings ($s)	13	11	13	12
Number of arrests	16	14	16	16
Early Head Start				
Bayley MDI scores	26	24	27	27
HOME scores	31	29	31	31
San Diego Cash-Out				
Value of purchased foods ($s per week)	25	24	28	22
Food energy as a percentage of the RDA	7	4	7	6
Teenage Parent Demonstration				
% of months in work/school activities	10	10	11	9
Monthly earnings ($)	10	9	10	11
Cash and Counseling				
All Medicaid expenditures ($)	53	53	62	52

Note: See Appendix Table B.1 in Schochet (2007) for study details and the text for variance formulas. Estimates for the finite-population model were made using the rank-order method for imputing missing potential outcomes.
[a]In this column, $p=0.5$ for the Job Corps Study where the actual p value is about two-thirds.

Table 4
Estimators for s_τ^2, $\boldsymbol{\tau}'\mathbf{P_X}\boldsymbol{\alpha}$, $\boldsymbol{\alpha}'\mathbf{P_X}\boldsymbol{\alpha}$ and $\boldsymbol{\tau}'\mathbf{P_X}\boldsymbol{\tau}$.

Evaluation and outcome measure	s_τ^2/s_{Pooled}^2 values, by imputation method		$(1/n)\boldsymbol{\tau}'\mathbf{P_X}\boldsymbol{\alpha}/s_{Pooled}^2$	$(1/n)\boldsymbol{\alpha}'\mathbf{P_X}\boldsymbol{\alpha}/s_{Pooled}^2$	$(1/n)\boldsymbol{\tau}'\mathbf{P_X}\boldsymbol{\tau}/s_{Pooled}^2$
	Subgroup	Rank-order	Rank-order method	Rank-order method	Rank-order method
NYC Vouchers					
Reading scores (percentile points)	0.06	0.01	0.01	0.41	0.00
Math scores (percentile points)	0.08	0.02	−0.02	0.33	0.00
Teach for America					
Math scores (percentile points)	0.01	0.19	−0.03	0.45	0.01
Reading scores (percentile points)	0.01	0.11	0.01	0.67	0.01
21st Century					
Reading scores (percentile points)	0.17	0.40	0.01	0.23	0.06
Math scores (percentile points)	0.33	0.40	0.00	0.17	0.08
Disciplinary problems index	0.14	0.11	0.00	0.08	0.03
Job Corps					
Year 4 earnings per week ($s)	0.02	0.11	0.01	0.11	0.01
Number of arrests	0.01	0.10	−0.01	0.14	0.00
Early Head Start					
Bayley MDI scores	0.15	0.21	−0.02	0.24	0.06
HOME scores	0.12	0.21	−0.02	0.28	0.06
San Diego Cash-Out					
Value of purchased foods ($s per Week)	0.04	0.09	0.00	0.23	0.02
Food energy as a percentage of the RDA	0.03	0.14	−0.01	0.04	0.01
Teenage Parent Demonstration					
% of months in work/school activities	0.03	0.04	0.00	0.09	0.00
Monthly earnings ($)	0.02	0.04	0.01	0.09	0.00
Cash and Counseling					
All Medicaid expenditures ($)	0.02	0.18	0.11	0.49	0.04

Note: See Appendix Table B.1 in Schochet (2007) for study details and the text for parameter definitions.

Table 5
s^2 and R^2 values by treatment status and white standard errors.

Evaluation and outcome measure	s^2 values		Regression R^2 values		Standard errors from models with covariates	
	Treatments (s_T^2)	Controls (s_C^2)	Treatments	Controls	Reduced-form	White
NYC Vouchers						
Reading scores (percentile points)	447.8	420.6+	0.40	0.44	0.998	0.959
Math scores (percentile points)	507.1	571.8+	0.27	0.43	1.178	1.112
Teach for America						
Math scores (percentile points)	314.7	347.1	0.48	0.52	0.653*	0.653*
Reading scores (percentile points)	410.6	432.8	0.66	0.72+	0.579	0.573
21st Century						
Reading scores (percentile points)	552.8	724.8+	0.30	0.23	1.159	1.129
Math scores (percentile points)	745.0	1,005.1+	0.25	0.21	1.685	1.635
Disciplinary problems index	0.7	0.7	0.12	0.11	0.064	0.038
Job Corps						
Year 4 earnings per week ($s)	39,636.0	35,603.9+	0.12	0.11	3.624*	3.595*
Number of arrests	1.9	2.3+	0.14	0.13	0.027*	0.026*
Early Head Start						
Bayley MDI scores	153.0	166.4	0.25	0.29	0.552*	0.545*
HOME scores	21.8	24.1+	0.29	0.33	0.202*	0.190*
San Diego Cash-Out						
Value of purchased foods ($s per Week)	1,641.4	1,949.4+	0.27	0.24	2.263*	2.200*
Food Energy as a Percentage of the RDA	3,289.5	4,457.6+	0.04	0.05	3.733**	3.691**
Teenage Parent Demonstration						
% of months in work/school activities	11.1	10.4	0.10	0.11	1.158*	1.157*
Monthly earnings ($)	3,684.9	3,213.4+	0.11	0.09	2.111*	2.093*
Cash and Counseling						
All Medicaid expenditures ($)	145,618.7	111,271.4+	0.58	0.48+	6.952*	6.961*

Note: See Appendix Table B.1 in Schochet (2007) for study details.
*ATE estimate is significantly different from zero at the 0.05 level, two-tailed test.
**ATE estimate is significantly different from zero at the 0.10 level, two-tailed test.
+The *F*-statistic to test for differences in s^2 values for treatments and controls is statistically significantly at the 0.05 level, or the difference in the R^2 values for treatments and controls is statistically significantly at the 0.05 level (based on bootstrap methods).

6. Conclusions

The theory presented in this paper suggests that the common practice of estimating ATEs using multivariate OLS regression models yields consistent and asymptotically normal ATE estimators under both the Neyman causal finite- and super-population models. OLS variance estimators, however, are supported by the Neyman model only under certain conditions. In particular, the variances are justified if subject-level treatment effects are constant, and are largely justified if the sample contains equal numbers of treatment and control subjects. Even if these conditions do not hold, however, efficiency gains from regression adjustment under the Neyman model will occur for most combinations of key variance parameters (such as the size and signs of the covariances between α_i, τ_i, and the covariates), although it is an empirical issue as to the specific parameter values that are likely to apply in practice.

Using data from eight large experiments testing the effectiveness of various social policy interventions in the United States, similar precision gains were found from regression adjustment under the OLS and Neyman models. This occurs primarily because treatment effects do not appear to vary substantially across subjects. These findings could be due to the relatively homogeneous, low-income populations included in each study and the relatively small overall impact estimates.

There are several lessons from this paper. First, analysts should first assess the extent of heterogeneity in subject-level treatment effects before applying OLS multivariate methods, and should compare ATE results using various estimation approaches for analyzing the data (such as estimating models without covariates, using variance estimators that allow for a flexible error structure, and including treatment-by-covariate interaction terms as explanatory variables). Second, theory suggests that the OLS framework is largely supported under stratified experimental designs. Thus, to the extent that causally-relevant blocking covariates can be found, stratification should be used liberally to obviate the need for ex-post regression adjustment. Finally, if treatment effects are expected to be heterogeneous and other design features (such as cost) are not an issue, there is a strong theoretical argument for setting the treatment group sampling rate at 50 percent, because under this design, precision gains from regression adjustment will occur under the Neyman causal model.

An important topic for future research is to empirically examine the justification for regression adjustment in experiments using data from other studies—such as drug trials. The empirical findings could differ from those found here for RCTs that target more diverse study populations and that find larger estimated treatment effects.

Appendix A. Proofs

Proof of Lemma 1. Applying (8) with $\tilde{\mathbf{Z}}$ as a column of 1 s yields $b_{SR,FP} = \bar{y}_T - \bar{y}_C$. To calculate the moments of $b_{SR,FP}$, we express $b_{SR,FP}$ as follows:

$$b_{SR,FP} = \frac{\sum_{i=1}^n (T_i - p)y_i}{np(1-p)} = \frac{\sum_{i=1}^n (T_i - p)[\beta_0 + \beta_{ATE}(T_i - p) + u_i]}{np(1-p)} = \beta_{ATE} + \frac{\sum_{i=1}^n (T_i - p)u_i}{np(1-p)}, \tag{A.1}$$

where the last equality holds because $\sum_i (T_i - p) = 0$ and $\sum_i (T_i - p)^2 = np(1-p)$. Substituting for u_i using (3b) and (3c) yields:

$$(b_{SR,FP} - \beta_{ATE}) = \frac{\sum_{i=1}^n [\alpha_i (T_i - p) + \tau_i (T_i - p)^2]}{np(1-p)} = \frac{\sum_{i=1}^n [\alpha_i + (1-2p)\tau_i] T_i}{np(1-p)}$$
$$= \frac{\sum_{i=1}^n l_i T_i}{np(1-p)}; \quad l_i = (1-p)(Y_{Ti} - \bar{Y}_T) + p(Y_{Ci} - \bar{Y}_C). \tag{A.2}$$

Thus, $E(b_{SR,FP} - \beta_{ATE}) = \sum_i l_i p / np(1-p) = 0$ because $\sum_i l_i = 0$. Thus, $b_{SR,FP}$ is unbiased.

Using (A.2), the variance of $b_{SR,FP}$ is

$$Var(b_{SR,FP}) = \frac{Var(\sum_{i=1}^n l_i T_i)}{[np(1-p)]^2} = \frac{p(1-p)\left(\sum_{i=1}^n l_i^2 - \dfrac{2}{(n-1)} \sum_{i=1}^n \sum_{j>i}^n l_i l_j\right)}{[np(1-p)]^2},$$

where the last equality holds because $Var(T_i) = p(1-p)$ and $Cov(T_i, T_j) = -p(1-p)/(n-1)$. Because $\sum_i l_i = 0$, it follows that $(\sum_i l_i)^2 = 0$, and thus, $-2\sum_i \sum_{j>i} l_i l_j = \sum_i l_i^2$. Hence,

$$Var(b_{SR,FP}) = \frac{\sum_i^n l_i^2}{np(1-p)(n-1)} = \frac{\sum_i^n [(1-p)(Y_{Ti} - \bar{Y}_T) + p(Y_{Ci} - \bar{Y}_C)]^2}{np(1-p)(n-1)}$$
$$= \frac{1}{np(1-p)} [(1-p)^2 S_T^2 + p^2 S_C^2 + 2p(1-p) S_{TC}^2], \tag{A.3}$$

where S_T^2 is the variance of the potential treatment outcomes across the n sample members, S_C^2 is the variance of the potential control outcomes, and $S_{TC}^2 = [1/(n-1)] \sum_{i=1}^n (Y_{Ti} - \bar{Y}_T)(Y_{Ci} - \bar{Y}_C)$ is the covariance between the treatment and control potential outcomes. A more conventional variance expression is obtained by writing S_τ^2, the variance of the subject-level treatment effects, as $S_\tau^2 = S_T^2 + S_C^2 - 2S_{TC}^2$. Solving for S_{TC}^2 and substituting into (A.3) yields the variance expression in (9). The asymptotic variance expression in (11) follows directly from (10).

The asymptotic normality of $b_{SR,FP}$ follows by expressing (A.1) as $\sqrt{n}p(1-p)(b_{SR,FP} - \beta_{ATE}) = \sum (T_i - p)u_i / \sqrt{n}$ and using a central limit theorem for finite populations (see for example, Freedman, 2008; Hogland, 1978; Hájek, 1960). □

Proof of Lemma 2. Applying (8) with $\tilde{Z}$ as a column of 1 s yields $b_{SR,SP} = \sum_{i=1}^n \tilde{T}_i y_i / np(1-p) = \bar{y}_T - \bar{y}_C$. Substituting for y_i using (5) yields $(b_{SR,SP} - \mu_{SP,\tau}) = \sum_{i=1}^n \tilde{T}_i \theta_i / np(1-p)$. Thus, $E(b_{SR,SP}) = \mu_{SP,\tau}$ because of (6b), and the variance expression in (12) is obtained using (6c) and (6d). Asymptotic normality follows by applying a conventional central limit theorem to $\sum \tilde{T}_i \theta_i / \sqrt{n}$ (see, for example, Rao, 1973). □

Proof of Lemma 3. The least squares estimator for $b_{MR,FP}$ is given by (8). If we substitute for $\mathbf{y}$ in (8) using the true model in (3), then $b_{MR,FP}$ can be expressed as follows:

$$b_{MR,FP} = \left[\frac{1}{n} \tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}})\tilde{\mathbf{T}}\right]^{-1} \frac{1}{n} \tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}})[\mathbf{K}\beta_0 + \tilde{\mathbf{T}}\beta_{ATE} + \mathbf{u}]$$
$$= \beta_{ATE} + \left[\frac{\tilde{\mathbf{T}}'(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}})\tilde{\mathbf{T}}}{n}\right]^{-1} \left[\frac{\tilde{\mathbf{T}}'\mathbf{u}}{n} - \frac{\tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\mathbf{u}}{n}\right]. \tag{A.4}$$

The estimator $b_{MR,FP}$ is biased in finite samples. However, we show that the bias tends to zero as n approaches infinity by examining the limiting values of each bracketed term:

$$\left[\frac{\tilde{\mathbf{T}}'(\mathbf{I}-\mathbf{P}_{\tilde{\mathbf{X}}})\tilde{\mathbf{T}}}{n}\right]^{-1} = \left[\frac{\tilde{\mathbf{T}}'\tilde{\mathbf{T}}}{n} - \left(\frac{\tilde{\mathbf{T}}'\tilde{\mathbf{X}}}{n}\right)\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n}\right)^{-1}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{T}}}{n}\right)\right]^{-1} \xrightarrow{p} \frac{1}{p(1-p)},$$

$$\frac{\tilde{\mathbf{T}}'\mathbf{u}}{n} = \frac{\sum_{i=1}^{n}(T_i-p)\alpha_i}{n} + \frac{\sum_{i=1}^{n}(T_i-p)^2\tau_i}{n} \xrightarrow{p} 0 + p(1-p)(0) = 0,$$

so that $\tilde{\mathbf{T}}$ and $\mathbf{u}$ are asymptotically uncorrelated, and:

$$\frac{\tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\mathbf{u}}{n} = \left(\frac{\tilde{\mathbf{T}}'\tilde{\mathbf{X}}}{n}\right)\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n}\right)^{-1}\left(\frac{\tilde{\mathbf{X}}'\mathbf{u}}{n}\right) \xrightarrow{p} (0)(\boldsymbol{\Omega}_{\mathbf{XX}}^{-1})(\boldsymbol{\Omega}_{\mathbf{X}\boldsymbol{\alpha}}) = 0,$$

where $\xrightarrow{p}$ denotes convergence in probability. Thus, $b_{MR,FP}$ is a consistent estimator.

To calculate the asymptotic variance of $b_{MR,FP}$, we apply an asymptotic expansion to (A.4):

$$b_{MR,FP} - \beta_{ATE} = \frac{\tilde{\mathbf{T}}'\mathbf{u}}{np(1-p)} - \frac{\tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}}{np(1-p)} + o_p(1/n), \tag{A.5}$$

where $o_p(1/n)$ signifies terms of order $1/n$. Note that the first term on the right-hand side of (A.5) pertains to the regression estimator without covariates. Note also that for the second term, $\tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} = \mathbf{T}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}$ because $-p\mathbf{K}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} = -\mathbf{pK}'\boldsymbol{\alpha} = 0$. Thus, (A.5) can be expressed as follows:

$$b_{MR,FP} - \beta_{ATE} = \frac{1}{np(1-p)}\sum_{i=1}^{n}[\alpha_i + (1-2p)\tau_i - \tilde{\mathbf{x}}_{\mathbf{i}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\boldsymbol{\alpha}]T_i + o_p(1/n), \tag{A.6}$$

where $\tilde{\mathbf{x}}_{\mathbf{i}}(1\times k)$ is a row vector of covariate values for subject i.

The term inside the brackets in (A.6) sums to zero because $\sum_i\alpha_i = \sum_i\tau_i = 0$, and $\sum_i\tilde{\mathbf{x}}_{\mathbf{i}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\boldsymbol{\alpha} = \sum_i\alpha_i = 0$ because it is the sum of fitted values when $\boldsymbol{\alpha}$ is regressed on $\tilde{\mathbf{X}}$. Thus, if we define l_i as the bracketed term in (A.6), then $\sum_i l_i = 0$, and we can use the same methods as for the regression estimator in Lemma 1 to derive the asymptotic variance of $b_{MR,FP}$:

$$\begin{aligned} Var(b_{MR,FP}) &= \frac{1}{np(1-p)}\frac{n}{n-1}\frac{\sum_{i=1}^{n}[\alpha_i + (1-2p)\tau_i - \tilde{\mathbf{x}}_{\mathbf{i}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\boldsymbol{\alpha}]^2}{n} + o_p(1/n) \\ &\xrightarrow{p} \left(\frac{\bar{S}_T^2}{np} + \frac{\bar{S}_C^2}{n(1-p)} - \frac{\bar{S}_\tau^2}{n}\right) - \frac{\boldsymbol{\Omega}_{\mathbf{X}\boldsymbol{\alpha}}'\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\boldsymbol{\Omega}_{\mathbf{X}\boldsymbol{\alpha}}}{np(1-p)} - 2(1-2p)\frac{\boldsymbol{\Omega}_{\mathbf{X}\boldsymbol{\tau}}'\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\boldsymbol{\Omega}_{\mathbf{X}\boldsymbol{\alpha}}}{np(1-p)}. \end{aligned}$$

The asymptotic normality of $b_{MR,FP}$ follows from (A.5) because both $\tilde{\mathbf{T}}'\mathbf{u}/\sqrt{n}$ and $\tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha}/\sqrt{n}$ are asymptotically normal (see Freedman, 2008; Hogland, 1978). □

Proof of Lemma 4. The least squares estimator for $b_{MR,SP}$ is given by (8). This estimator is consistent because $E(\tilde{T}_iX_{ij}^*) = E(\tilde{T}_i\theta_i) = 0$, and thus, $(b_{MR,SP} - \mu_{ATE}) \xrightarrow{p} \{E[\tilde{\mathbf{T}}'(\mathbf{I}-\mathbf{P}_{\mathbf{X}^*})\tilde{\mathbf{T}}]\}^{-1}E[\tilde{\mathbf{T}}'(\mathbf{I}-\mathbf{P}_{\mathbf{X}^*})\boldsymbol{\theta}] = 0$. The asymptotic variance of $b_{MR,SP}$ is obtained by using a standard asymptotic expansion to express the estimator as $(b_{MR,SP} - \mu_{ATE}) = (np(1-p))^{-1}[\tilde{\mathbf{T}}'\boldsymbol{\theta} - \tilde{\mathbf{T}}'\mathbf{P}_{\mathbf{X}^*}\boldsymbol{\alpha}] + o_p(1/n)$, and then calculating $[np(1-p)]^{-2}\sum_{i=1}^{n}E[\tilde{T}_i(\alpha_i + \tilde{T}_i\tau_i - \mathbf{x}_{\mathbf{i}}^*\boldsymbol{\Lambda}_{\mathbf{XX}}^{-1}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}})]^2$, which after some algebra yields (16). The asymptotic normality of $b_{MR,SP}$ follows using a standard central limit theorem. □

Proof of Lemma 5. The least squares estimator for $b_{MR,SP,STR}$ is given by (8). To show that this estimator is consistent, we note first that as n becomes large $(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}}/n)^{-1}$ converges to a block diagonal matrix, because the off-diagonal terms, $E(\tilde{T}_iX_{ij}^*)$, $E(\tilde{T}_i^2X_{ij}^*)$, and $E(\tilde{T}_iX_{ij}^{*2})$ are zero due to the centering of the variables. Thus, $\mathbf{P}_{\tilde{\mathbf{Z}}} \xrightarrow{p} E(\mathbf{P}_{\tilde{\mathbf{K}}} + \mathbf{P}_{\mathbf{X}^*} + \mathbf{P}_{\mathbf{Q}^*})$ where $Q_{ij}^* = X_{ij}^*\tilde{T}_i$ and hence,

$$b_{MR,SP,STR} - \mu_{ATE} \xrightarrow{p} \{E[\tilde{\mathbf{T}}'(\mathbf{I}-\mathbf{P}_{\mathbf{X}^*}-\mathbf{P}_{\mathbf{Q}^*})\tilde{\mathbf{T}}]\}^{-1}E[\tilde{\mathbf{T}}'(\mathbf{I}-\mathbf{P}_{\mathbf{X}^*}-\mathbf{P}_{\mathbf{Q}^*})\boldsymbol{\theta}] = 0$$

To calculate the asymptotic variance of $b_{MR,SP,STR}$, we use a standard asymptotic expansion to express $b_{MR,SP,STR}$ as follows:

$$b_{MR,SP,STR} - \mu_{ATE} = (np(1-p))^{-1}[\tilde{\mathbf{T}}'\boldsymbol{\theta} - \tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{X}}}\boldsymbol{\alpha} - \tilde{\mathbf{T}}'\mathbf{P}_{\tilde{\mathbf{Q}}}\boldsymbol{\alpha}] + o_p(1/n)$$

Note that $(\tilde{\mathbf{Q}}'\tilde{\mathbf{Q}}/n)^{-1}(\tilde{\mathbf{Q}}'\boldsymbol{\alpha}/n) \xrightarrow{p} \boldsymbol{\Lambda}_{\mathbf{XX}}^{-1}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\tau}}$. Thus, the asymptotic variance of $b_{MR,SP,STR}$ can be obtained by expanding the following expression:

$$[np(1-p)]^{-2}\sum_{i=1}^{n}\{E[\tilde{T}_i(\alpha_i+\tilde{T}_i\tau_i-\mathbf{x_i^*}\boldsymbol{\Lambda}_{\mathbf{XX}}^{-1}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\alpha}}-\tilde{T}_i\mathbf{x_i^*}\boldsymbol{\Lambda}_{\mathbf{XX}}^{-\mathbf{1}}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\tau}})]^2+E[p(1-p)\mathbf{x_i^*}\boldsymbol{\Lambda}_{\mathbf{XX}}^{-\mathbf{1}}\boldsymbol{\Lambda}_{\mathbf{X}\boldsymbol{\tau}}]^2\}$$

which after some algebra yields (18). Asymptotic normality follows using a standard central limit theorem. □

References

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70 (1), 91–117.

Byrk, A., Raudenbush, S., 1992. Hierarchical Linear Models for Social and Behavioral Research. Applications and Data Analysis Methods. Sage, Newbury Park, CA.

Cochran, W., Cox, G., 1957. Experimental Designs. Wiley, New York.

Dale, S., Brown, R., 2005. The Effect of Cash and Counseling on Medicaid and Medicare Costs: Findings for Adults in Three States. Mathematica Policy Research, Princeton, NJ.

Decker, P., Glazerman, S., Mayer, D., 2004. The Effects of Teach For America on Students: Findings from a National Evaluation. Mathematica Policy Research, Princeton, NJ.

Dynarski, M., Moore, M., Rosenberg, L., James-Burdumy, S., Deke, J., Mansfield, W., 2004. When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, New Findings. Mathematica Policy Research, Princeton, NJ.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. Wiley, New York.

Freedman, D., 2008. On regression adjustments to experimental data. Advances in Applied Mathematics 40, 180–193.

Hájek, J., 1960. Limiting distributions in simple random sampling from a finite population. Magyar Tudoanyos Akademia Budapest Matematikai Kutato Intezet Koezlemenyei 5, 361–374 (Trudy Publications).

Heckman, J., Smith, J., Clements, N., 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. Review of Economic Studies 64 (4), 487–535.

Heckman, J., 2001. Micro data, heterogeneity, and the evaluation of public policy: nobel lecture. Journal of Political Economy 109 (4), 673–748.

Heckman, J., 2005. The scientific model of causality. Sociological Methodology 35 (1), 1–97.

Heckman, J., Vytlacil, E., 2006. Econometric evaluation of social programs. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6. Elsevier, Amsterdam.

Hogland, T., 1978. Sampling from a finite population: a remainder term estimate. Scandinavian Journal of Statistics 5, 69–71.

Holland, P., 1986. Statistics and causal inference. Journal of the American Statistical Association 81 (396), 945–960.

Imbens, G., Rubin, D., 2001. Neyman's Repeated Sampling Perspective in Completely Randomized Experiments. Unpublished Manuscript, Harvard University Statistics Department (Chapter 5).

Love, J., et al., 2002. Building Their Futures: How Early Head Start Programs are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Mathematica Policy Research, Princeton, NJ.

Mayer, D., Peterson, P., Myers, D., Tuttle, C., Howell, W., 2002. School Choice in New York City: An Evaluation of the School Choice Scholarships Program. Mathematica Policy Research, Washington, DC.

Maynard, R., Nicholson, W., Rangarajan, A., 1993. Breaking the Cycle of Poverty: The Effectiveness of Mandatory Services for Welfare-Dependent Teenage Parents. Mathematica Policy Research, Princeton, NJ.

Neyman, J., 1923. On the Application of Probability Theory to Agricultural Experiments: Essay on Principles. Translated in Statistical Science 5(4), 465–472 (Section 9, 1990).

Ohls, J., Fraker, T., Martini, A., Ponza, M., 1992. The Effects of Cash-Out on Food Use by Food Stamp Program Participants in San Diego. Mathematica, Princeton, NJ.

Rao, C., 1973. Linear Statistical Inference and Its Applications. Wiley, New York.

Rubin, D., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Education Psychology 66, 688–701.

Rubin, D., 1978. Bayesian inference for causal effects: The role of randomization. Annals of statistics 6 (1), 34–58.

Schochet, P., 2007. Is Regression Adjustment Supported by the Neyman Model for Causal Inference?. Mathematica Policy Research, Princeton, NJ. (Working Paper).

Schochet, P., Burghardt, J., McConnell, S., 2008. Does Job Corps Work? Impact Findings from the National Job Corps Study. American Economic Review 98 (5), 1864–1886.

Smith, J. Whalley, A.,Wilcox, N., 2005. Are Program Participants Good Evaluators? Working Paper, University of Michigan.

White, H., 1980. A heteroscedasticity-consistent covariance estimator and a direct test for heteroscedasticity. Econometrica 48, 817–883.

Yang, L., Tsiatis, A., 2001. Efficiency study of estimators for a treatment effects in a pretest–posttest trial. American Statistician 55 (4), 314–321.