SPC

International Journal of Advanced Statistics and Probability, 2 (2) (2014) 56-69
©Science Publishing Corporation
www.sciencepubco.com/index.php/IJASP
doi: 10.14419/ijasp.v2i2.2895
Research Paper

Neyman's causal model with stochastic potential outcomes: implications for the completely randomized design

Emil Scosyrev Ph.D.

University of Rochester 601 Elmwood Ave Rochester, New York, USA 14642
E-mail: ens79ur@gmail.com

Copyright © 2014 Scosyrev. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In Neyman's causal model (NCM), each subject participating in a two-arm randomized trial has a pair of potential outcomes – one outcome would be observed under treatment and another under control. In the stochastic version of NCM the two potential outcomes are viewed as possibly non-degenerate random variables with finite expectations and variances. The subject-level treatment effect is the expected outcome under treatment minus that under control, and the average treatment effect is the arithmetic mean of the subject-level effects. In the present paper properties of the ordinary "difference of means" estimator and its associated variance estimator are examined in the completely randomized design with stochastic potential outcomes. Estimation theory is developed under randomization distribution without commitment to any particular probability model for enrollment, because in real trials subjects are not enrolled by a sampling mechanism with known selection probabilities. It is shown that in this theoretical framework, the "difference of means" estimator is asymptotically normal and consistent for the average treatment effect in the study cohort, while its associated variance estimator is conservative, producing confidence intervals with at least nominal asymptotic coverage. The proofs are not trivial because in the randomization framework sample means under treatment and control are correlated random variables.

Keywords: *Causality; Clinical Trials; Internal Validity; Neyman's Causal Model; Randomization-Based Inference; Stochastic Potential Outcomes.*

1. Introduction

The purpose of randomized trials is to make inference about treatment effects. In the original version of Neyman's causal model (NCM), a treatment effect in a particular set (Ω) of N subjects was defined as follows [1-3]. Let i represent the physical identity of a given subject, and let Y_{iT} and Y_{iC} denote the potential outcomes of the i th subject under treatment and under control intervention, respectively. If the i th subject is treated, the observed outcome is Y_{iT} , while under control it is Y_{iC} . These potential outcomes, although not jointly observable, were defined as real numbers having objective simultaneous existence even prior to treatment assignment. The subject-level treatment effect is $\delta_i = Y_{iT} - Y_{iC}$, and the average treatment effect in Ω is the arithmetic mean of the subject-level effects

$$\delta = N^{-1}\sum_{i=1}^{N}\delta_i = N^{-1}\sum_{i=1}^{N}Y_{iT} - N^{-1}\sum_{i=1}^{N}Y_{iC}$$

Because the concept of potential outcomes as simultaneously existing real numbers does not allow for stochastic effects of post-interventional factors such as random measurement error, Neyman [4] proposed a more general framework with stochastic potential outcomes. In the stochastic NCM, Y_{iT} and Y_{iC} are viewed as possibly non-degenerate random variables with finite but otherwise unconstrained expectations (y_{iT} , y_{iC}) and variances (υ_{iT} , υ_{iC}). If the i th subject is treated, the observed outcome Y_i is a realization of Y_{iT} , while under control it is a realization of Y_{iC} .The subject-level causal effect is $\delta_i = y_{iT} - y_{iC}$. If everyone in Ω is treated, the expected mean outcome is

$$\mu_T = N^{-1}\sum_{i=1}^{N} y_{iT}$$

Similarly, if everyone in Ω receives the control intervention, the expected mean outcome is

$$\mu_C = N^{-1}\sum_{i=1}^{N} y_{iC}$$

The average treatment effect in Ω is

$$\delta = N^{-1}\sum_{i=1}^{N} \delta_i = \mu_T - \mu_C$$

It should be noted that stochastic NCM contains deterministic potential outcomes as a special case ($\upsilon_{iT} = 0, \upsilon_{iC} = 0$ for all i). Hence, any estimator with desirable statistical properties in the stochastic framework will have these properties in the deterministic framework, while the converse is not necessarily true (i.e., the stochastic framework makes weaker assumptions).

For causal inference in randomized experiments it is important to make a distinction between internal and external validity of estimation. Suppose that Ω denotes the set of all subjects who participated in the trial, and let δ denote the average treatment effect in Ω. Similarly, let Ω' denote the set of all subjects who were eligible for inclusion in the trial during the accrual period ($\Omega \subset \Omega'$) and let δ' denote the average treatment effect in Ω'. We will say that a point estimator of the average treatment effect is internally valid if it is consistent for δ under randomization distribution, and is externally valid if it is consistent for δ' in the process of accrual followed by randomization. Similarly, an interval estimator is internally valid if it has at least nominal coverage for δ under randomization distribution, and is externally valid if it has at least nominal coverage for δ' in the process of accrual followed by randomization.

To establish internal validity, properties of estimators must be examined under randomization distribution without a probability model for enrollment. In contrast, for external validity these properties must be studied in the process of accrual followed by randomization. The latter is straightforward when subjects are selected from Ω' by a sampling mechanism with known selection probabilities, such as simple random sampling, cluster sampling, stratified sampling, etc. However, in most clinical trials subjects are not enrolled by a mechanism with known selection probabilities. Hence, external validity usually depends on assumptions about accrual process that are not fully verifiable. Given these considerations, it can be recommended that any estimator developed under an assumed probability model for enrollment should also have a demonstrated property of internal validity under randomization distribution. An internally valid estimate of δ can be used as a reasonable guess about average treatment effect in patients similar to those actually enrolled in the trial, whether or not these patients happen to be representative of the entire population of eligible subjects.

It should be noted that formal proofs of internal validity of a given interval estimator may be somewhat challenging because randomization induces correlation between functions of subject-level outcomes observed under alternative interventions. For example, arithmetic means of subject-level outcomes in the treatment and the control arms are correlated under randomization distribution [5]. These randomization-induced correlations must be taken into account in the mathematical formalism for proof of internal validity of a given interval estimator [5], [6]. In particular, because treatment and control means are correlated random variables, the variance of their difference is not the sum of individual variances, and the standard proofs of the Central Limit Theorem (CLT) for the "difference of means" are not applicable.

In the present paper the framework of stochastic NCM is applied to the completely randomized design to establish internal validity of one of the most frequently used interval estimators. Let T and C denote the sets of subjects assigned to treatment and to control, respectively, and define for each intervention arm the means of subject-specific observed outcomes

$$\hat{\mu}_T = n_T^{-1}\sum_{i\in T} Y_i \ , \ \hat{\mu}_C = n_C^{-1}\sum_{i\in C} Y_i$$

Let $\hat{\delta} = \hat{\mu}_T - \hat{\mu}_C$. The ordinary "sum of variances" estimator is

$$\hat{V}(\hat{\delta}) = \frac{(n_T - 1)^{-1}\sum_{i\in T}(Y_i - \hat{\mu}_T)^2}{n_T} + \frac{(n_C - 1)^{-1}\sum_{i\in C}(Y_i - \hat{\mu}_C)^2}{n_C} \tag{1}$$

A well-known interval estimator of δ is $\hat{\delta} \pm z_{\alpha/2}\hat{V}(\hat{\delta})^{1/2}$. Our objective is to show that under completely randomized treatment assignment, when N is not too small, this interval estimator has at least nominal coverage probability for δ. The rest of this paper is organized as follows. Mathematical theory is presented in Section 2. A numerical example illustrating the proposed theory is considered in Section 3. Section 4 contains discussion of the main points of the paper.

2. Mathematical theory

In the completely randomized design we take from Ω a simple random sample of n_T subjects and allocate them to the treatment arm, while the remaining n_C subjects are allocated to the control arm. With this randomization mechanism, n_T and n_C are fixed by design. If subjects are enrolled consecutively (one after another) as is usually the case in clinical trials, the completely randomized design can be implemented by creating a randomization plan as follows. Take from the set of the first N natural numbers $(1, 2 \ldots N)$ a simple random sample (S) of n_T numbers and assign subjects whose order of enrollment is in S to treatment, with the remaining subjects assigned to control.

To prove internal validity of $\hat{\delta} \pm z_{\alpha/2} \hat{V}(\hat{\delta})^{1/2}$ in this data-generating process, we will need to show that $\hat{\delta}$ is consistent and asymptotically normal and $\hat{V}(\hat{\delta})$ is conservative. To state these results more formally and obtain general proofs, define on Ω the following finite population means, variances, and covariance of the subject-level parameters

$$E(\upsilon_{iT}) = N^{-1}\sum_{i=1}^{N}\upsilon_{iT} \ , \ E(\upsilon_{iC}) = N^{-1}\sum_{i=1}^{N}\upsilon_{iC}$$

$$V(y_{iT}) = N^{-1}\sum_{i=1}^{N}(y_{iT} - \mu_T)^2 \ , \ V(y_{iC}) = N^{-1}\sum_{i=1}^{N}(y_{iC} - \mu_C)^2$$

$$Cov(y_{iT}, y_{iC}) = N^{-1}\sum_{i=1}^{N}(y_{iT} - \mu_T)(y_{iC} - \mu_C)$$

Also define for the treatment and the control arms the means of subject-specific expected outcomes

$$m_T = n_T^{-1}\sum_{i \in T} y_{iT} \ , \ m_C = n_C^{-1}\sum_{i \in C} y_{iC}$$

Given that the r th set of subjects is randomized to treatment, responses $Y_i, \forall i \in T$ are assumed to be independent, although not necessarily identically distributed. This conditional independence assumption is highly plausible when the outcome of one subject does not influence the outcomes of the other subjects (e.g., as is usually the case in medical trials with non-contagious outcomes). Given that the r th set of subjects is randomized to treatment, expectation of $\hat{\mu}_T$ in these subjects is

$$E(\hat{\mu}_T \mid r) = n_T^{-1}E(\sum_{i \in T} Y_i \mid r) = n_T^{-1}\sum_{i \in T} y_{iT} = m_T$$

Marginally with respect to treatment assignment, by the law of the iterated mean $E(\hat{\mu}_T) = E[E(\hat{\mu}_T \mid r)] = E(m_T) = \mu_T$. This last equality follows from the fact that T is a simple random sample from Ω and hence m_T, which is the sample mean of y_{iT} in T is unbiased for the population mean of y_{iT} in Ω, which is μ_T. Of course, by the same argument, $E(\hat{\mu}_C) = \mu_C$. Because expectation of a difference is a difference of expectations (even for correlated random variables), the expectation of $\hat{\delta} = (\hat{\mu}_T - \hat{\mu}_C)$ is equal to δ. In other words, in this data-generating process, the difference of sample means is unbiased for the true treatment effect.

We now consider the variance of $\hat{\delta}$. Again, given that the r th set of subjects is randomized to treatment, responses $Y_i, \forall i \in T$ are independent, and the variance of $\hat{\mu}_T$ in these subjects is

$$V(\hat{\mu}_T \mid r) = n_T^{-2}[\sum_{i \in T} V(Y_i) \mid r] = (n_T^{-1}\sum_{i \in T} \upsilon_{iT}) / n_T$$

Since $n_T^{-1}\sum_{i \in T} \upsilon_{iT}$ is a sample mean of υ_{iT} in T, its expectation marginally with respect to treatment assignment is the population mean $N^{-1}\sum_{i=1}^{N}\upsilon_{iT}$. Hence,

$$E[V(\hat{\mu}_T \mid r)] = E(n_T^{-1}\sum_{i \in T} \upsilon_{iT}) / n_T = (N^{-1}\sum_{i=1}^{N}\upsilon_{iT}) / n_T$$

Now, by the law of variance decomposition, marginally with respect to treatment assignment

$$V(\hat{\mu}_T) = E[V(\hat{\mu}_T \mid r)] + V[E(\hat{\mu}_T \mid r)] = n_T^{-1}(N^{-1}\sum_{i=1}^{N}\upsilon_{iT}) + V(m_T) = \frac{E(\upsilon_{iT})}{n_T} + \frac{N - n_T}{N - 1}\frac{V(y_{iT})}{n_T} \quad (2)$$

In the last expression, $V(m_T)$ is just the variance of the sample mean of y_{iT} in the process of simple random sampling from a finite population (Ω), and $(N - n_T)/(N - 1)$ is the finite population correction, well-known from elementary sampling theory. Of course, same arguments apply to the control set. Hence,

$$V(\hat{\mu}_C) = \frac{E(\upsilon_{iC})}{n_C} + \frac{N - n_C}{N - 1}\frac{V(y_{iC})}{n_C} \tag{3}$$

Obviously, the variance of $\hat{\delta}$ can be written as $V(\hat{\mu}_T - \hat{\mu}_C) = V(\hat{\mu}_T) + V(\hat{\mu}_C) - 2Cov(\hat{\mu}_T, \hat{\mu}_C)$. The covariance term turns out to be equal to

$$Cov(\hat{\mu}_T, \hat{\mu}_C) = -Cov(y_{iT}, y_{iC}) / (N - 1) \tag{4}$$

Neyman [1], [2] proved this result for the case of deterministic potential outcomes (i.e., $\upsilon_{iT} = \upsilon_{iC} = 0$ for all i) assuming equal allocation ($n_T = n_C$). However, equation (4) also holds in the more general formulation with stochastic potential outcomes and equal or unequal allocation. The general proof of (4) is given in Appendix 1. It follows that

$$V(\hat{\delta}) = V(\hat{\mu}_T) + V(\hat{\mu}_C) - 2Cov(\hat{\mu}_T, \hat{\mu}_C) =$$

$$= \frac{E(\upsilon_{iT})}{n_T} + \frac{N - n_T}{N - 1}\frac{V(y_{iT})}{n_T} + \frac{E(\upsilon_{iC})}{n_C} + \frac{N - n_C}{N - 1}\frac{V(y_{iC})}{n_C} + \frac{2Cov(y_{iT}, y_{iC})}{N - 1} \tag{5}$$

It is seen that for any given $p_T = n_T / N$, $p_C = n_C / N$, $E(\upsilon_{iT})$, $E(\upsilon_{iC})$, $V(y_{iT})$, $V(y_{iC})$ and $Cov(y_{iT}, y_{iC})$, expression (5) goes to zero with increasing N. Hence, $\hat{\delta}$ is consistent for δ. However, (5) itself cannot be estimated consistently because $E(\upsilon_{iT})$ and $V(y_{iT})$ cannot be estimated separately and likewise for $E(\upsilon_{iC})$ and $V(y_{iC})$, and $Cov(y_{iT}, y_{iC})$ is not identifiable. Nevertheless, it will be shown below that $\hat{V}(\hat{\delta})$ is a conservative estimator of $V(\hat{\delta})$, $E[\hat{V}(\hat{\delta})] \geq V(\hat{\delta})$.

Expectation of $\hat{V}(\hat{\delta})$ is the sum of expectations of the two terms in (1). It is shown in Appendix 2 that these expectations are

$$E\left[\frac{(n_T - 1)^{-1}\sum_{i \in T}(Y_i - \hat{\mu}_T)^2}{n_T}\right] = \frac{E(\upsilon_{iT})}{n_T} + \frac{N}{N - 1}\frac{V(y_{iT})}{n_T} \tag{6}$$

$$E\left[\frac{(n_C - 1)^{-1}\sum_{i \in C}(Y_i - \hat{\mu}_C)^2}{n_C}\right] = \frac{E(\upsilon_{iC})}{n_C} + \frac{N}{N - 1}\frac{V(y_{iC})}{n_C} \tag{7}$$

$$E[\hat{V}(\hat{\delta})] = \frac{E(\upsilon_{iT})}{n_T} + \frac{N}{N - 1}\frac{V(y_{iT})}{n_T} + \frac{E(\upsilon_{iC})}{n_C} + \frac{N}{N - 1}\frac{V(y_{iC})}{n_C} \tag{8}$$

To prove that $E[\hat{V}(\hat{\delta})] \geq V(\hat{\delta})$ for any joint frequency distribution of $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ in Ω, consider the ratio $V(\hat{\delta}) / E[\hat{V}(\hat{\delta})]$, which is (5) divided by (8). In this ratio, the numerator (5) takes its maximum value for $Corr(y_{iT}, y_{iC}) = 1$ when $Cov(y_{iT}, y_{iC}) = \sqrt{V(y_{iT})}\sqrt{V(y_{iC})}$. Hence, to prove that (1) is unbiased or conservative for the true variance in (5), it is sufficient to verify that

$$\frac{E(\upsilon_{iT})}{n_T} + \frac{N - n_T}{N - 1}\frac{V(y_{iT})}{n_T} + \frac{E(\upsilon_{iC})}{n_C} + \frac{N - n_C}{N - 1}\frac{V(y_{iC})}{n_C} + 2\frac{\sqrt{V(y_{iT})}\sqrt{V(y_{iC})}}{N - 1} \leq$$
$$\frac{E(\upsilon_{iT})}{n_T} + \frac{N}{N - 1}\frac{V(y_{iT})}{n_T} + \frac{E(\upsilon_{iC})}{n_C} + \frac{N}{N - 1}\frac{V(y_{iC})}{n_C}$$

After subtracting from both sides of this inequality $E(\upsilon_{iT}) / n_T$ and $E(\upsilon_{iC}) / n_C$ and multiplying both sides by $(N - 1) / N$, we arrive at

$$\frac{N - n_T}{N}\frac{V(y_{iT})}{n_T} + \frac{N - n_C}{N}\frac{V(y_{iC})}{n_C} + 2\frac{\sqrt{V(y_{iT})}\sqrt{V(y_{iC})}}{N} \leq \frac{V(y_{iT})}{n_T} + \frac{V(y_{iC})}{n_C} \tag{9}$$

Subtracting the left side of (9) from both sides, we obtain

$$\frac{V(y_{iT})}{N} + \frac{V(y_{iC})}{N} - 2\frac{\sqrt{V(y_{iT})}\sqrt{V(y_{iC})}}{N} \geq 0$$
$$\left[\sqrt{\frac{V(y_{iT})}{N}} - \sqrt{\frac{V(y_{iC})}{N}}\right]^2 \geq 0$$

This proves that the standard variance estimator (1) is unbiased or conservative given any joint distribution of $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ in Ω. It is worth emphasizing that for given p_T, p_C, $E(\upsilon_{iT})$, $E(\upsilon_{iC})$, $V(y_{iT})$, $V(y_{iC})$ and $Cov(y_{iT}, y_{iC})$, the ratio of (5) to (8) does not go to unity with increasing N. In fact, as $N \to \infty$, $N-1 \to N$, and

$$\frac{V(\hat{\delta})}{E[\hat{V}(\delta)]} \to \frac{p_T^{-1}[E(\upsilon_{iT}) + p_C V(y_{iT})] + p_C^{-1}[E(\upsilon_{iC}) + p_T V(y_{iC})] + 2Cov(y_{iT}, y_{iC})}{p_T^{-1}[E(\upsilon_{iT}) + V(y_{iT})] + p_C^{-1}[E(\upsilon_{iC}) + V(y_{iC})]} \tag{10}$$

Expression (10) does not depend on N, and as we already showed, cannot exceed unity. Because $V(\hat{\delta})/\hat{V}(\hat{\delta})$ also goes to (10) in probability (see Appendix 2), it follows from Slutsky's Theorem [7] that confidence intervals constructed based on (1) in the usual manner as $\hat{\delta} \pm z_{\alpha/2}\hat{V}(\hat{\delta})^{1/2}$ have at least nominal asymptotic coverage, and the corresponding Wald-type hypotheses tests asymptotically do not exceed the nominal alpha-level, provided that $\hat{\delta}$ is asymptotically normal. The classical proofs of the CLT are not applicable in the randomization framework due to the correlation of $\hat{\mu}_T$ and $\hat{\mu}_C$, which does not go to zero with increasing N, as shown in Appendix 1. However, Freedman [6] proved asymptotic normality of $\hat{\delta}$ in the completely randomized design with deterministic potential outcomes. A more general proof applicable to our current stochastic formulation is given in Appendix 3. It is shown that for large N $(\hat{\delta} - \delta)/\hat{V}(\hat{\delta})^{1/2}$ is approximately normally distributed with mean zero and variance given by (10). Because this variance cannot exceed unity, we have for large N approximately $\Pr[-z_{1-\alpha/2} \le (\hat{\delta} - \delta)/\hat{V}(\hat{\delta})^{1/2} \le z_{1-\alpha/2}] \ge 1-\alpha$, which can be used for hypothesis testing, and $\Pr[\hat{\delta} - z_{1-\alpha/2}\hat{V}(\hat{\delta})^{1/2} \le \delta \le \hat{\delta} + z_{1-\alpha/2}\hat{V}(\hat{\delta})^{1/2}] \ge 1-\alpha$, which is a statement of coverage probability of nominally $(1-\alpha)\times 100\%$ confidence intervals. The actual coverage depends on individual terms in (10).

3. Numerical example

The aim of this section is to illustrate the main theoretical results derived in Section 2. For this purpose, consider a set of 4 subjects in Table 1. Each subject can receive either treatment or control. Suppose that the outcome of interest is a binary variable, representing for example vital status at 5 years of follow-up (1=alive, 0=dead). Because the outcomes are binary, expectations y_{iT} and y_{iC} represent the 5-year survival probabilities under treatment and under control, respectively. Also, for binary outcomes, $\upsilon_{iT} = y_{iT}(1-y_{iT})$, $\upsilon_{iC} = y_{iC}(1-y_{iC})$. If all 4 subjects are treated, their average 5-year survival probability is $\mu_T = 0.6$, while under control it is $\mu_C = 0.25$. The average treatment effect is $\delta = 0.6 - 0.25 = 0.35$.

Table 1 Expectations and Variances of Binary Stochastic Potential Outcomes for a Set of Four Subjects

Subject	y_{iT}	y_{iC}	υ_{iT}	υ_{iC}
1	0.9	0.3	0.09	0.21
2	0.8	0.2	0.16	0.16
3	0.1	0.1	0.09	0.09
4	0.6	0.4	0.24	0.24
Mean	0.60	0.25	0.145	0.175
Var	0.095	0.0125		

$Cov(y_{iT}, y_{iC}) = 0.02$

Suppose that in order to estimate this treatment effect, we take a simple random sample of 2 subjects and allocate them to treatment, while the remaining 2 subjects are allocated to control. Because we only have 4 subjects and the observed outcomes are binary, the sample space for this data-generating process contains only 6 possible treatment assignments with 16 possible realizations given each assignment, which are written down explicitly in Table 2. The probability of each realization given a particular treatment assignment can be computed by noting that individual responses are independent; hence their joint probability is the product of marginal probabilities.

For example, given that subjects 1 and 2 are assigned to treatment and subjects 3 and 4 are assigned to control (T/C = 1,2/3,4), the probability that each of the four subjects will be alive at 5 years (i.e., R1 in Table 2) is $y_{1T} \times y_{2T} \times y_{3C} \times y_{4C} = 0.9 \times 0.8 \times 0.1 \times 0.4 = 0.0288$, and the marginal probability of this assignment and outcome combination is 0.0288/6 (we divide by 6 because there are 6 equally likely treatment assignments). The second realization (R2) in Table 2 under the same assignment (T/C = 1,2/3,4) differs from the first only in that subject 4 did not survive, so the conditional probability of that realization given that particular treatment assignment is $y_{1T} \times y_{2T} \times y_{3C} \times (1-y_{4C}) = 0.9 \times 0.8 \times 0.1 \times 0.6 = 0.0432$, and the marginal probability of this assignment and outcome combination is 0.0432/6. The probabilities of all other realizations are computed based on the same reasoning.

Table 2 Simulated Example: Sample Space with Associated Probabilities

	Observed outcomes								Pr(R) given T / C					
R	Y_T	Y_T '	Y_C	Y_C '	$\hat{\mu}_T$	$\hat{\mu}_C$	$\hat{\delta}$	$\hat{V}(\hat{\delta})$	1,2/3,4	1,3/2,4	1,4/2,3	2,3/1,4	2,4/1,3	3,4/1,2
1	1	1	1	1	1	1	0	0	0.0288	0.0072	0.0108	0.0096	0.0144	0.0036
2	1	1	1	0	1	0.5	0.5	0.25	0.0432	0.0108	0.0972	0.0144	0.1296	0.0144
3	1	1	0	1	1	0.5	0.5	0.25	0.2592	0.0288	0.0432	0.0224	0.0336	0.0084
4	1	1	0	0	1	0	1	0	0.3888	0.0432	0.3888	0.0336	0.3024	0.0336
5	1	0	1	1	0.5	1	-0.5	0.25	0.0072	0.0648	0.0072	0.0864	0.0096	0.0024
6	1	0	1	0	0.5	0.5	0	0.5	0.0108	0.0972	0.0648	0.1296	0.0864	0.0096
7	1	0	0	1	0.5	0.5	0	0.5	0.0648	0.2592	0.0288	0.2016	0.0224	0.0056
8	1	0	0	0	0.5	0	0.5	0.25	0.0972	0.3888	0.2592	0.3024	0.2016	0.0224
9	0	1	1	1	0.5	1	-0.5	0.25	0.0032	0.0008	0.0012	0.0024	0.0036	0.0324
10	0	1	1	0	0.5	0.5	0	0.5	0.0048	0.0012	0.0108	0.0036	0.0324	0.1296
11	0	1	0	1	0.5	0.5	0	0.5	0.0288	0.0032	0.0048	0.0056	0.0084	0.0756
12	0	1	0	0	0.5	0	0.5	0.25	0.0432	0.0048	0.0432	0.0084	0.0756	0.3024
13	0	0	1	1	0	1	-1	0	0.0008	0.0072	0.0008	0.0216	0.0024	0.0216
14	0	0	1	0	0	0.5	-0.5	0.25	0.0012	0.0108	0.0072	0.0324	0.0216	0.0864
15	0	0	0	1	0	0.5	-0.5	0.25	0.0072	0.0288	0.0032	0.0504	0.0056	0.0504
16	0	0	0	0	0	0	0	0	0.0108	0.0432	0.0288	0.0756	0.0504	0.2016

R = realization; Y_T and Y_T ' are the observed outcomes of the two subjects assigned to treatment, Y_C and Y_C ' are the observed outcomes of the two subjects assigned to control; T is the set of subjects assigned to treatment, C is the set of subjects assigned to control.

We can now verify the key equations in Section 2. The theory in Section 2 predicts that $E(\hat{\mu}_T) = 0.6$, $E(\hat{\mu}_C) = 0.25$, and $E(\hat{\delta}) = 0.35$. This can be verified by computing these expectations directly from the sample space (Table 2)

$$E(\hat{\mu}_T) = (1\times 0.0288 + 1\times 0.0432 + ... + 0\times 0.2016)/6 = 0.6$$
$$E(\hat{\mu}_C) = (1\times 0.0288 + 0.5\times 0.0432 + ... + 0\times 0.2016)/6 = 0.25$$
$$E(\hat{\delta}) = (0\times 0.0288 + 0.5\times 0.0432 + ... + 0\times 0.2016)/6 = 0.35$$

Similarly, to verify the variance and covariance expressions (2)-(5), and expectation of $\hat{V}(\hat{\delta})$ in (8) we may compute directly from the sample space

$$V(\hat{\mu}_T) = [(1-0.6)^2\times 0.0288 + (1-0.6)^2\times 0.0432 + ... + (0-0.6)^2\times 0.2016]/6 = 0.3125/3$$
$$V(\hat{\mu}_C) = [(1-0.25)^2\times 0.0288 + (0.5-0.25)^2\times 0.0432 + ... + (0-0.25)^2\times 0.2016]/6 = 0.275/3$$
$$Cov(\hat{\mu}_T, \hat{\mu}_C) = [(1-0.6)(1-0.25)\times 0.0288 + (1-0.6)(0.5-0.25)\times 0.0432 + ...$$
$$... + (0-0.6)(0-0.25)\times 0.2016]/6 = -0.02/3$$
$$V(\hat{\delta}) = [(0-0.35)^2\times 0.0288 + (0.5-0.35)^2\times 0.0432 + ... + (0-0.35)^2\times 0.2016]/6 = 0.6275/3$$
$$E[\hat{V}(\hat{\delta})] = (0\times 0.0288 + 0.25\times 0.0432 + ... + 0\times 0.2016)/6 = 0.695/3$$

Same results can be obtained by applying (2)-(5) and (8) to data in Table 1

$$V(\hat{\mu}_T) = \frac{0.145}{2} + (\frac{4-2}{4-1})\frac{0.095}{2} = \frac{0.3125}{3}$$
$$V(\hat{\mu}_C) = \frac{0.175}{2} + (\frac{4-2}{4-1})\frac{0.0125}{2} = \frac{0.275}{3}$$
$$Cov(\hat{\mu}_T, \hat{\mu}_C) = -0.02/3$$
$$V(\hat{\delta}) = \frac{0.145}{2} + (\frac{4-2}{4-1})\frac{0.095}{2} + \frac{0.175}{2} + (\frac{4-2}{4-1})\frac{0.0125}{2} + \frac{2\times 0.02}{4-1} = \frac{0.6275}{3}$$
$$E[\hat{V}(\hat{\delta})] = \frac{0.145}{2} + (\frac{4}{4-1})\frac{0.095}{2} + \frac{0.175}{2} + (\frac{4}{4-1})\frac{0.0125}{2} = \frac{0.695}{3}$$

For illustration of asymptotic results, the following simulation study was performed in SAS 9.3. A total of 50 copies of the dataset in Table 1 were created, resulting in a set of 200 subjects. From this set, a simple random sample of 100 subjects was obtained and assigned to treatment, while the remaining 100 subjects were assigned to control. The treatment effect was estimated as in Section 2 by the difference of means $(\hat{\delta})$, and the variance of $\hat{\delta}$ was estimated by (1). This process was repeated 1,000 times. Of course, as before $\delta = 0.35$. From (5), $V(\hat{\delta}) = 0.0039$ and from (8) $E[\hat{V}(\hat{\delta})] = 0.0043$. Based on 1,000 randomizations, the observed frequency distribution of $\hat{\delta}$ was nearly normal, with mean 0.3499 and variance 0.0038. The observed mean of $\hat{V}(\hat{\delta})$ was 0.0043 and the observed coverage of 95% confidence intervals $\hat{\delta} \pm 1.96\hat{V}(\hat{\delta})^{1/2}$ was 97.0%.

4. Discussion

Randomization-based inference with stochastic potential outcomes was introduced by Neyman [4] in the context of agricultural experiments, where he considered a comparison of s interventions in a randomized experiment performed on a field of N plots (subjects). The yield of each plot under a given intervention was assumed to be a normal random variable with constant variance taking the same value for all plots under all interventions. This is similar to our current formalism, except that the assumptions of normality and constant variance were not made here (e.g., the subject-level responses may even be Bernoulli variables, as in Section 3). The average treatment effect (δ) in Neyman [4] was defined for any two interventions as in our Section 2. To estimate δ, Neyman [4] assumed that the N plots were grouped in n' blocks of s plots per block, where s is also the number of interventions being compared. Hence, the number of plots per block was constrained to be equal to the number of interventions. With two interventions (treatment and control), there are two plots per block. Randomization was performed within each block by randomly distributing the s plots to the s interventions, one plot per intervention. This was repeated for each block. Neyman [4] referred to this process as the method of randomized blocks.

Because in Neyman's [4] analysis, the number of plots per block was constrained to be the same as the number of interventions being compared, his mathematical formalism cannot be used to describe a completely randomized design. In particular, we cannot obtain expectation or variance expressions for $\hat{\delta}$ in a two-arm completely randomized trial by simply setting the number of blocks in Neyman's [4] equations to one, because this would imply in his mathematical formalism that only two subjects are being randomized. In a two-arm completely randomized trial with more than 2 subjects, Neyman's [4] equations no longer apply, which was the main motivation for the present paper. In a two-arm trial with more than 2 subjects, Neyman's equations describe a block-randomized design, but with the assumption of only two subjects per block, and with restrictions on the variance components as already described. Since these assumptions are too restrictive for trials involving human subjects, it may be appropriate to generalize Neyman's [4] equations to block-randomized designs with these assumptions removed. This may be a subject of further research. It should also be noted that Neyman [4] did not study the asymptotic distribution of $\hat{\delta}$. A proof of the CLT for $\hat{\delta}$ under completely randomized assignment with stochastic potential outcomes is given in Appendix 3 of the present paper.

Historically, most work on Neyman's causal model was done under the assumption of deterministic potential outcomes [6, 8-10]. This formalism goes back to Neyman's earlier paper [1], [2]. For recent work on this formalism see Freedman [6], Schochet [8], Lin [9], and Miratrix et al [10]. Freedman [6] established results for interval estimation under completely randomized treatment assignment similar to those obtained in the present paper but with the assumption of deterministic potential outcomes. Freedman [6] also showed that in the completely randomized design the ordinary least-squares variance estimator from the linear model fitted with or without covariates does not have internal validity (i.e., may produce anticonservative confidence intervals). Lin [9] showed that this problem is corrected by the robust (sandwich) variance estimator. Schochet [8] studied internal validity of certain estimators in the cluster randomized design, while Miratrix [10] considered block-randomized design and post-stratification.

To what extent findings on internal validity of estimators obtained under the assumption of deterministic potential outcomes [6, 8-10] hold in the more general stochastic framework is not immediately obvious. It is certainly possible for a given interval estimator to be conservative in the deterministic framework and become anticonservative with stochastic potential outcomes [11]. In fact, stochastic generalization may even invalidate certain point estimators of the average treatment effect. For example, the classical "instrumental variables" (IV) estimator that can be used to adjust for non-compliance in clinical trials depends on the assumption of simultaneously existing compliance states defined under assignment to alternative interventions [12]. When this assumption is relaxed by allowing for compliance probabilities, the IV argument loses its generality, and the IV estimator becomes inconsistent for any well-defined treatment effect except under certain special conditions (such as full compliance in one intervention arm) [12]. Because the choice between deterministic and the more general stochastic frameworks has practical implications for the choice of estimators, it is important to discuss the physical meaning of the subject-level variance components $\{\upsilon_{iT}, \upsilon_{iC}\}$.

The purpose of the $\{\upsilon_{iT}, \upsilon_{iC}\}$ terms is to allow for possible dependence of potential outcomes on uncontrolled initial conditions under which the assigned interventions are actually administered and to represent the stochastic effects of post-interventional factors, such as random measurement error. For illustration of this point, consider the following example. Patients with low and intermediate risk prostate cancer are often managed with active surveillance, which involves among other things random biopsies of the prostate gland (e.g., once every two years) to assess for evidence of disease progression. A biopsy is performed by taking a predetermined number (usually 12) of tissue specimens ("cores") from the prostate gland for microscopic examination. The biopsies are random because cancerous regions in the gland usually cannot be distinguished from normal tissue until the specimens are examined under the microscope. The proportion of positive cores (i.e., the number of cores with prostate cancer divided by the total number of cores) is an important indicator of disease progression. Suppose that we perform a randomized trial of a new dietary supplement (versus placebo) believed to reduce the risk of prostate cancer progression in patients on active surveillance, with the proportion of positive cores at 2 years from randomization as one of the end-points. If the i th subject is treated, the proportion of positive cores is Y_{iT}, while under control it is Y_{iC}.

The framework with deterministic potential outcomes assumes that both Y_{iT} and Y_{iC} have definite values prior to randomization. In reality however Y_{iT} and Y_{iC} are sample proportions and hence cannot take definite values until the random biopsy sample has been collected (even if there are no other stochastic effects of post-interventional factors). This is recognized in the stochastic NCM, where prior to the actual measurement of the outcome, Y_{iT} and Y_{iC} are viewed as possibly non-degenerate random variables with expectations $\{y_{iT}, y_{iC}\}$ and variances $\{\upsilon_{iT}, \upsilon_{iC}\}$. If the i th subject is treated, the observed proportion of positive cores (Y_i) is a realization of Y_{iT}, while under control it would be a realization of Y_{iC} .The subject-level parameters $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ are assumed to take finite values, but are otherwise unconstrained. The subject-level treatment effect is the expected proportion of positive cores under treatment (y_{iT}) minus that under control (y_{iC}).

In this example, the true quantity that we may wish to know is the proportion of the total gland volume involved by cancer, but because this cannot be measured without taking the entire gland out, we resort to an estimate based on the random biopsy technique. The difference between the observed proportion of positive cores and the proportion of total gland volume involved by cancer may be thought of as a random measurement error. This is allowed by stochastic NCM but is not allowed by the framework with deterministic potential outcomes. It is important to emphasize however that because stochastic NCM contains deterministic potential outcomes as a special case ($\upsilon_{iT} = \upsilon_{iC} = 0 \forall i \in \Omega$), any estimator with desirable statistical properties in the stochastic framework will have these properties in the deterministic framework, while the converse is not always true. Hence, the stochastic framework is safer because it makes weaker assumptions about physical reality of randomized experiments.

Another important feature of the theory developed in this paper is its focus on internal validity of the trial (i.e., valid causal inference for the set of subjects undergoing randomization). Because in real trials subjects are not enrolled by any sampling mechanism with known selection probabilities, external validity of a given estimator is not guaranteed by randomization alone or by any other statistical considerations. One way to formally examine this problem is to let p_j denote the (unknown) probability of enrollment in the trial for a given subject in Ω'. Although it would be convenient to assume that these enrollment probabilities are the same for all subjects in Ω', this is usually not a realistic assumption. Hence, we must let p_j vary between the subjects. Let π denote the mean of p_j in Ω' and define the relative selection probability $r_j = p_j / \pi$. It can be shown [13] that under arbitrary unequal probability sampling mechanism followed by completely randomized treatment assignment, the expectation of $\hat{\delta}$ is equal to δ' plus the covariance of the subject-level treatment effect with the subject-level relative selection probability in Ω'

$$E(\hat{\delta}) = \delta' + Cov(\delta_j, r_j)$$

In other words, the difference of means estimator is unbiased for δ' if and only if the subject-level treatment effects are uncorrelated with the subject-level relative selection probabilities. Plausibility of this condition may be assessed based on expert opinion; however, this condition is not guaranteed by any statistical considerations unless participants are enrolled in the trial from a sampling frame with known selection probabilities. It should also be noted that although condition $Cov(\delta_j, r_j) = 0$ (holding at least approximately) is necessary for external validity of the standard interval estimator based on (1), it is not sufficient because it does not guarantee that (1) is consistent or conservative for the true variance of $\hat{\delta}$ in the process of accrual followed by randomization, nor is the CLT for $\hat{\delta}$ guaranteed to hold.

If we let $i = 1,...,N$ count the order of enrollment of the subjects in the trial, then with respect to accrual process each set of subject-level parameters $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ may be viewed as a random vector. If accrual process is such that $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ are independent and identically distributed (iid) random vectors and $Cov(\delta_j, r_j) = 0$, then in the process of accrual followed by randomization $\hat{\delta}$ is unbiased for δ', (1) is consistent, and interval estimator based on (1) has nominal asymptotic coverage for δ' [3]. The assumption of identical distributions for random vectors $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ is plausible when the N subjects are drawn from the same large population of eligible subjects (defined by the same inclusion criteria), and the joint relative frequency distribution of subject-level parameters in this population remains stable during the study. The assumption of independence for $\{y_{iT}, y_{iC}, \upsilon_{iT}, \upsilon_{iC}\}$ is plausible when the sampling units are individual subjects rather than clusters of subjects with similar potential outcomes, and one subject's decision to participate does not influence other subject's decisions. The assumption of no correlation between subject-level treatment effects and subject-level selection probabilities is plausible when the set of the trial participants is similar to the population of eligible subjects with respect to the distribution of known or suspected effect modifiers (e.g., age, stage of disease, etc.). Although all of the above assumptions can be assessed for plausibility based on expert opinion and understanding of accrual process in any given trial, they are not guaranteed to hold by randomization or by any other feature of the study design. In contrast, as was shown in the present paper, internal validity can be ensured by randomization alone, provided there are no major protocol violations and randomization-induced correlations are properly taken into account. In general, it seems reasonable to require that any estimator developed under an assumed probability model for enrollment should also have the property of internal validity, as established in the present paper for $\hat{\delta} \pm z_{\alpha/2} \hat{V}(\hat{\delta})^{1/2}$.

5. Conclusions

In this paper, properties of the "difference of means" estimator, and its associated variance estimator were examined in the completely randomized design with stochastic potential outcomes. Estimation theory was developed under randomization distribution without a probability model for enrollment, because in real trials subjects are not enrolled by a sampling mechanism with known selection probabilities. It was shown that in this theoretical framework, the "difference of means" estimator is asymptotically normal and consistent for the average treatment effect in the study cohort, while the standard variance estimator is conservative, producing confidence intervals with at least nominal asymptotic coverage. To what extent conventional estimators used with other experimental designs (e.g., block-randomized or cluster randomized trials) retain their internal validity in stochastic NCM remains to be determined.

Acknowledgments

The author would like to thank his wife Tatyana for independently verifying all numerical results reported in Section 3.

References

[1] Neyman, J, Sur les applications de la th´eorie des probabilit´es aux experiences agricoles: Essai des principes, Roczniki Nauk Rolniczych. 10 (1923) 1–51, in Polish.
[2] Dabrowska, D.M., Speed T.P., On the application of probability theory to agricultural experiments: essay on principles, Statistical Science 5 (1990) 465-472. (English translation of the manuscript originally published by J Neyman in 1923 in Roczniki Nauk Rolniczych Tom X, 1-51).
[3] Rubin, D., Comment: Neyman (1923) and causal inference in experiments and observational studies. Statistical Science 5 (1990) 472-480.
[4] Neyman, J., Iwaszkiewicz K., Kolodziejczyk S., Statistical problems in agricultural experimentation, Journal of the Royal Statistical Society 2 (1935) 107-108.
[5] Scosyrev E., Interval estimation of treatment effects in randomized trials: when do confidence intervals have nominal coverage? International Statistical Review 80 (2012) 439-451.
[6] Freedman, D.A., On regression adjustments to experimental data, Advances in Applied Mathematics 40 (2008)180-193.
[7] Casella, G., Berger, R.L., Statistical Inference, Pacific Grove, Duxbury, 2000.
[8] Schochet P.Z., Estimators for clustered education RCTs using the Neyman model for causal inference, Journal of Educational and Behavioral Statistics 38 (2013) 219-238.
[9] Lin W., Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique, Annals of Applied Statistics 7 (2013) 295-318.
[10] Miratrix LW, Sekhon JS, Yu B., Adjusting treatment effect estimates by post-stratification in randomized experiments, Journal of the Royal Statistical Society 75 (2013) 369-396.
[11] Robins JM., Confidence intervals for causal parameters, Statistics in Medicine 7 (1988) 773-785.
[12] Scosyrev E., Identification of causal effects using instrumental variables in randomized trials with stochastic compliance, Biometrical Journal 55 (2013) 97-113.
[13] Scosyrev E., Estimation of population mean under unequal probability sampling with unknown selection probabilities, American Journal of Theoretical and Applied Statistics 3 (2014) 65-72.
[14] Cochran WG, Sampling Techniques. 3rd ed, Wiley & Sons, New York, 1977, p39-40.
[15] Parzen, E., Modern Probability Theory and Its Applications, Wiley & Sons, New York, 1960, p431.
[16] Johnson, N.L., Kotz S., Balakrishnan N., Continuous Univariate Distributions, Volume 1, Wiley & Sons, New York, 1994, p163.

Appendix 1

In this Appendix, we derive the expressions for $Cov(\hat{\mu}_T, \hat{\mu}_C)$ and $Corr(\hat{\mu}_T, \hat{\mu}_C)$ in the completely randomized design. First note that given a particular treatment assignment (r), treatment and control means $\hat{\mu}_T$ and $\hat{\mu}_C$ are independent random variables. Hence, $E(\hat{\mu}_T \hat{\mu}_C \mid r) = [E(\hat{\mu}_T)E(\hat{\mu}_C) \mid r] = m_T m_C$. Marginally with respect to treatment assignment

$$Cov(\hat{\mu}_T, \hat{\mu}_C) = E[(\hat{\mu}_T - \mu_T)(\hat{\mu}_C - \mu_C)] = E(\hat{\mu}_T \hat{\mu}_C) - \mu_T \mu_C =$$
$$= E[E(\hat{\mu}_T \hat{\mu}_C \mid r)] - \mu_T \mu_C = E(m_T m_C) - \mu_T \mu_C = Cov(m_T, m_C)$$

Hence, we find $Cov(\hat{\mu}_T, \hat{\mu}_C)$ if we find $E(m_T m_C)$. For this purpose, it is convenient to introduce indicator variables $t_i = 1$ if $i \in T$ (i.e., if the i th subject is assigned to treatment), $t_i = 0$ otherwise, and $c_i = 1$ if $i \in C$, $c_i = 0$ otherwise. With this notation $m_T = n_T^{-1} \sum_{i=1}^{N} y_{iT} t_i$ and $m_C = n_C^{-1} \sum_{i=1}^{N} y_{iC} c_i$, where only t_i and c_i are random variables. Now we can write

$$E(m_T m_C) = E[(n_T^{-1} \sum_{i=1}^{N} y_{iT} t_i)(n_C^{-1} \sum_{i=1}^{N} y_{iC} c_i)] =$$
$$= n_T^{-1} n_C^{-1} E[(y_{1T} t_1 + y_{2T} t_2 + ... + y_{NT} t_N)(y_{1C} c_1 + y_{2C} c_2 + ... + y_{NC} c_N)] =$$
$$= n_T^{-1} n_C^{-1} E[(\sum_{i=1}^{N} y_{iT} y_{iC} t_i c_i) + (\sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC} t_i c_j)]$$

In the last expression, $\sum_{i=1}^{N} y_{iT} y_{iC} t_i c_i$ is the sum of products based on the same subjects, while $\sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC} t_i c_j$ is the sum of products based on different subjects. The expectation of $\sum_{i=1}^{N} y_{iT} y_{iC} t_i c_i$ is clearly zero, because for each subject either t_i or c_i must be zero, hence their product is also zero. To find expectation of $\sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC} t_i c_j$, we need to find $E(t_i c_j)$ for $i \neq j$. Note that the product $t_i c_j$ is a binary variable taking the value of 1 if and only if both t_i and c_j are equal to 1. Hence, $E(t_i c_j) = \Pr(t_i = 1, c_j = 1)$, which is just the probability that for any two subjects i and j in Ω, subject i will be randomized to treatment and subject j will be randomized to control. With this notation, $\Pr(c_j = 1) = n_C N^{-1}$, $\Pr(t_i = 1 | c_j = 1) = n_T (N-1)^{-1}$ and $E(t_i c_j) = \Pr(t_i = 1, c_j = 1) = \Pr(t_i = 1 | c_j = 1) \Pr(c_j = 1) = n_T n_C N^{-1} (N-1)^{-1}$. It follows that $E(\sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC} t_i c_j) = n_T n_C N^{-1} (N-1)^{-1} \sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC}$, and hence

$$E(m_T m_C) = N^{-1} (N-1)^{-1} \sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC}$$

In the last expression, $\sum_{i=1}^{N} \sum_{j \neq i} y_{iT} y_{jC}$ is the sum of products based on different subjects, which can be written as the sum of products based on all subjects $(\sum_{i=1}^{N} y_{iT})(\sum_{i=1}^{N} y_{iC})$ minus the sum of products based on the same subjects $\sum_{i=1}^{N} y_{iT} y_{iC}$. Hence,

$$E(m_T m_C) = \frac{(\sum_{i=1}^{N} y_{iT})(\sum_{i=1}^{N} y_{iC}) - \sum_{i=1}^{N} y_{iT} y_{iC}}{N(N-1)} = \frac{N \mu_T N \mu_C - \sum_{i=1}^{N} y_{iT} y_{iC}}{N(N-1)}$$

Dividing the numerator and the denominator of the last equality by N, we obtain

$$E(m_T m_C) = \frac{N \mu_T \mu_C - N^{-1} \sum_{i=1}^{N} y_{iT} y_{iC}}{N-1} = \frac{(N-1)\mu_T \mu_C + \mu_T \mu_C - N^{-1} \sum_{i=1}^{N} y_{iT} y_{iC}}{N-1} =$$

$$= \mu_T \mu_C + \frac{(N^{-1} \sum_{i=1}^{N} y_{iT})(N^{-1} \sum_{i=1}^{N} y_{iC}) - N^{-1} \sum_{i=1}^{N} y_{iT} y_{iC}}{N-1} = \mu_T \mu_C - \frac{Cov(y_{iT}, y_{iC})}{N-1}$$

The last equality follows from the fact that the mean product is the product of means plus the covariance, that is $N^{-1} \sum_{i=1}^{N} y_{iT} y_{iC} = (N^{-1} \sum_{i=1}^{N} y_{iT})(N^{-1} \sum_{i=1}^{N} y_{iC}) + Cov(y_{iT}, y_{iC})$. Now

$$Cov(\hat{\mu}_T, \hat{\mu}_C) = Cov(m_T, m_C) = E(m_T m_C) - \mu_T \mu_C = \frac{-Cov(y_{iT}, y_{iC})}{N-1}$$

It is important to note that although for given p_T, p_C, $E(\upsilon_{iT})$, $E(\upsilon_{iC})$, $V(y_{iT})$, $V(y_{iC})$, and $Cov(y_{iT}, y_{iC})$, the covariance of $\hat{\mu}_T$ and $\hat{\mu}_C$ goes to zero with increasing N, $Corr(\hat{\mu}_T, \hat{\mu}_C)$ does not. To see this, substitute (2)-(4) in $Corr(\hat{\mu}_T, \hat{\mu}_C) = Cov(\hat{\mu}_T, \hat{\mu}_C) / \sqrt{V(\hat{\mu}_T)} \sqrt{V(\hat{\mu}_C)}$

$$Corr(\hat{\mu}_T, \hat{\mu}_C) = \frac{-Cov(y_{iT}, y_{iC}) / (N-1)}{\sqrt{[\frac{E(\upsilon_{iT})}{n_T} + \frac{N - n_T}{N-1} \frac{V(y_{iT})}{n_T}][\frac{E(\upsilon_{iC})}{n_C} + \frac{N - n_C}{N-1} \frac{V(y_{iC})}{n_C}]}} =$$

$$= \frac{-Cov(y_{iT}, y_{iC}) / (N-1)}{\sqrt{[\frac{E(\upsilon_{iT})}{p_T N} + \frac{p_C}{p_T} \frac{V(y_{iT})}{N-1}][\frac{E(\upsilon_{iC})}{p_C N} + \frac{p_T}{p_C} \frac{V(y_{iC})}{N-1}]}} =$$

$$= \frac{-Cov(y_{iT}, y_{iC}) / (N-1)}{\sqrt{[(\frac{N-1}{N-1}) \frac{E(\upsilon_{iT})}{p_T N} + \frac{p_C}{p_T} \frac{V(y_{iT})}{N-1}][(\frac{N-1}{N-1}) \frac{E(\upsilon_{iC})}{p_C N} + \frac{p_T}{p_C} \frac{V(y_{iC})}{N-1}]}} =$$

$$= \frac{-Cov(y_{iT}, y_{iC})}{\sqrt{(\frac{N-1}{N})^2 \frac{E(\upsilon_{iT}) E(\upsilon_{iC})}{p_T p_C} + (\frac{N-1}{N}) \frac{E(\upsilon_{iT}) V(y_{iC})}{p_C} + (\frac{N-1}{N}) \frac{V(y_{iT}) E(\upsilon_{iC})}{p_T} + \frac{V(y_{iT}) V(y_{iC})}{1}}}$$

Because $(N-1)/N \to 1$, as $N \to \infty$, $Corr(\hat{\mu}_T, \hat{\mu}_C)$ does not go to zero with increasing N, and $\hat{\mu}_T$ and $\hat{\mu}_C$ cannot be viewed as being approximately independent or even uncorrelated in any asymptotic arguments. Note that with deterministic potential outcomes $\hat{\mu}_T = m_T$, $\hat{\mu}_C = m_C$ and $Corr(\hat{\mu}_T, \hat{\mu}_C) = Corr(m_T, m_C) = -Corr(y_{iT}, y_{iC})$.

Appendix 2

In this Appendix we derive expressions (6) and (7), which in sum give $E[\hat{V}(\hat{\delta})]$ for the completely randomized design. Explicit proof of (6) is given below, while the proof of (7) is its mirror image (just replace T by C). Equation (8) is the sum of (6) and (7). It is also shown that $\hat{V}(\hat{\delta})/E_a[\hat{V}(\hat{\delta})] \xrightarrow{P} 1$, where $E_a[\hat{V}(\hat{\delta})] = [n_T^{-1}E(\upsilon_{iT}) + n_T^{-1}V(y_{iT}) + n_C^{-1}E(\upsilon_{iC}) + n_C^{-1}V(y_{iC})]$ is the asymptotic form of (8), and "P" denotes convergence in probability.

Recall that any variance expression can be written as the mean square minus the squared mean. This result will be used repeatedly in this Appendix. Now to prove equation (6), we first find $E[n_T^{-1}\Sigma_{i\in T}(Y_i - \hat{\mu}_T)^2] = E[(n_T^{-1}\Sigma_{i\in T}Y_i^2) - (n_T^{-1}\Sigma_{i\in T}Y_i)^2]$, where $E(n_T^{-1}\Sigma_{i\in T}Y_i^2)$ will be evaluated using the law of the iterated mean, and $E[(n_T^{-1}\Sigma_{i\in T}Y_i)^2]$ will be evaluated using (2)

$$E(n_T^{-1}\Sigma_{i\in T}Y_i^2 \mid r) = n_T^{-1}\Sigma_{i\in T}E(Y_i^2) \mid r = n_T^{-1}\Sigma_{i\in T}(\upsilon_{iT} + y_{iT}^2) = n_T^{-1}\Sigma_{i\in T}\upsilon_{iT} + n_T^{-1}\Sigma_{i\in T}y_{iT}^2$$

$$E(n_T^{-1}\Sigma_{i\in T}Y_i^2) = E(n_T^{-1}\Sigma_{i\in T}\upsilon_{iT}) + E(n_T^{-1}\Sigma_{i\in T}y_{iT}^2) = N^{-1}\Sigma_{i=1}^{N}\upsilon_{iT} + N^{-1}\Sigma_{i=1}^{N}y_{iT}^2 \quad \text{(i)}$$

$$E[(n_T^{-1}\Sigma_{i\in T}Y_i)^2] = V(n_T^{-1}\Sigma_{i\in T}Y_i) + [E(n_T^{-1}\Sigma_{i\in T}Y_i)]^2 = V(\hat{\mu}_T) + \mu_T^2 =$$

$$= \frac{E(\upsilon_{iT})}{n_T} + \frac{N-n_T}{N-1}\frac{V(y_{iT})}{n_T} + N^{-1}\Sigma_{i=1}^{N}y_{iT}^2 - V(y_{iT}) \quad \text{(ii)}$$

Now

$$E[n_T^{-1}\Sigma_{i\in T}(Y_i - \hat{\mu}_T)^2] = E(n_T^{-1}\Sigma_{i\in T}Y_i^2) - E[(n_T^{-1}\Sigma_{i\in T}Y_i)^2] =$$

$$= E(\upsilon_{iT}) + N^{-1}\Sigma_{i=1}^{N}y_{iT}^2 - \frac{E(\upsilon_{iT})}{n_T} - \frac{N-n_T}{N-1}\frac{V(y_{iT})}{n_T} - N^{-1}\Sigma_{i=1}^{N}y_{iT}^2 + V(y_{iT}) =$$

$$= \frac{n_T E(\upsilon_{iT}) - E(\upsilon_{iT})}{n_T} - \frac{N-n_T}{N-1}\frac{V(y_{iT})}{n_T} + \frac{(N-1)n_T V(y_{iT})}{(N-1)n_T} =$$

$$= \frac{E(\upsilon_{iT})(n_T-1)}{n_T} + \frac{Nn_T V(y_{iT}) - n_T V(y_{iT}) - NV(y_{iT}) + n_T V(y_{iT})}{(N-1)n_T} =$$

$$= \frac{E(\upsilon_{iT})(n_T-1)}{n_T} + \frac{N}{N-1}\frac{V(y_{iT})(n_T-1)}{n_T}$$

Hence,

$$E\left[\frac{(n_T-1)^{-1}\Sigma_{i\in T}(Y_i - \hat{\mu}_T)^2}{n_T}\right] = \frac{E(\upsilon_{iT})}{n_T} + \frac{N}{N-1}\frac{V(y_{iT})}{n_T}$$

Expression (7) is proved by a similar argument; while (8) is the sum of (6) and (7). Now we will show that $\hat{V}(\hat{\delta})/E_a[\hat{V}(\hat{\delta})] \xrightarrow{P} 1$ as $N \to \infty$. To state this result more precisely, we need to define the limiting process by writing down some regularity conditions, that is, by specifying what attributes of Ω are to be held fixed as $N \to \infty$. For this purpose, we need to introduce some additional notation. Let $\{y'_{iT}, \upsilon'_{iT}\}$ and $\{y'_{iC}, \upsilon'_{iC}\}$ denote means and variances of the squared subject-level outcomes (Y_i^2) under treatment and under control intervention respectively. Note that $\{y'_{iT}, \upsilon'_{iT}\}$ and $\{y'_{iC}, \upsilon'_{iC}\}$ are parameters of the subject-level distributions, just like subject parameters $\{y_{iT}, \upsilon_{iT}\}$ and $\{y_{iC}, \upsilon_{iC}\}$. Define $V(y'_{iT}) = N^{-1}\Sigma_{i=1}^{N}(y'_{iT} - N^{-1}\Sigma_{i=1}^{N}y'_{iT})^2$, $E(\upsilon'_{iT}) = N^{-1}\Sigma_{i=1}^{N}\upsilon'_{iT}$, $V(y'_{iC}) = N^{-1}\Sigma_{i=1}^{N}(y'_{iC} - N^{-1}\Sigma_{i=1}^{N}y'_{iC})^2$, $E(\upsilon'_{iC}) = N^{-1}\Sigma_{i=1}^{N}\upsilon'_{iC}$. Also, let γ_{iT} and γ_{iC} denote the fourth central moments of the subject-level distributions of the i th's subject under treatment and under control, respectively. That is, if the i th subject is treated, then $E[|Y_i - y_{iT}|^4] = \gamma_{iT}$, while under control $E[|Y_i - y_{iC}|^4] = \gamma_{iC}$. Let $\{\gamma_{iT}^{\max}, \gamma_{iT}^{\min}\}$ and $\{\gamma_{iC}^{\max}, \gamma_{iC}^{\min}\}$ be the maximum and the minimum values of γ_{iT} and γ_{iC} in Ω, respectively, and use similar notation for y_{iT}, y_{iC}, υ_{iT}, υ_{iC}. Now we are ready to state the regularity conditions for the limiting process

Regularity conditions: As $N \to \infty$, let p_T, p_C, μ_T, μ_C, $V(y_{iT})$, $V(y_{iC})$, $Cov(y_{iT}, y_{iC})$, $E(\upsilon_{iT})$, $E(\upsilon_{iC})$, $V(y'_{iT})$, $V(y'_{iC})$, $E(\upsilon'_{iT})$, $E(\upsilon'_{iC})$, $\{y_{iT}^{\max}, y_{iT}^{\min}\}$, $\{y_{iC}^{\max}, y_{iC}^{\min}\}$, $\{\gamma_{iT}^{\max}, \gamma_{iT}^{\min}\}$, $\{\gamma_{iC}^{\max}, \gamma_{iC}^{\min}\}$, $\{\upsilon_{iT}^{\max}, \upsilon_{iT}^{\min}\}$, $\{\upsilon_{iC}^{\max}, \upsilon_{iC}^{\min}\}$ remain constant, while other attributes of Ω may or may not change.

Given these regularity conditions, note that since $(n_T - 1) \to n_T$, $(n_C - 1) \to n_C$, as $N \to \infty$, and because $n_T = p_T N$, $n_C = p_C N$, we have from (1) and (8) as $N \to \infty$

$$\frac{\hat{V}(\hat{\delta})}{E_a[\hat{V}(\hat{\delta})]} \to \frac{p_T^{-1}n_T^{-1}\Sigma_{i\in T}(Y_i-\hat{\mu}_T)^2 + p_C^{-1}n_C^{-1}\Sigma_{i\in C}(Y_i-\hat{\mu}_C)^2}{p_T^{-1}[E(\upsilon_{iT})+V(y_{iT})]+p_C^{-1}[E(\upsilon_{iC})+V(y_{iC})]} \quad \text{(iii)}$$

Recall that in general, for two sequences of random variables x_1 and x_2, and two constants k_1 and k_2, if $x_1 \xrightarrow{P} k_1$ and $x_2 \xrightarrow{P} k_2$, then $x_1k_2 \xrightarrow{P} k_1k_2$, $x_1^2 \xrightarrow{P} k_1^2$, $(x_1+x_2) \xrightarrow{P} (k_1+k_2)$, and $(x_1+x_2)/(k_1+k_2) \xrightarrow{P} 1$ [7]. Hence, to prove that (iii) goes in probability to unity, we just need to show that (a) $n_T^{-1}\Sigma_{i\in T}(Y_i-\hat{\mu}_T)^2 \xrightarrow{P} E(\upsilon_{iT})+V(y_{iT})$, and (b) $n_C^{-1}\Sigma_{i\in C}(Y_i-\hat{\mu}_C)^2 \xrightarrow{P} E(\upsilon_{iC})+V(y_{iC})$. Explicit proof of (a) is given below, while the proof of (b) is its mirror image (just replace T by C).

To prove (a), we may again write $n_T^{-1}\Sigma_{i\in T}(Y_i-\hat{\mu}_T)^2$ as $(n_T^{-1}\Sigma_{i\in T}Y_i^2)-(n_T^{-1}\Sigma_{i\in T}Y_i)^2$ and show that the first term $(n_T^{-1}\Sigma_{i\in T}Y_i^2)$ is consistent for its expectation (i), while the second term $(n_T^{-1}\Sigma_{i\in T}Y_i)^2$ is consistent for μ_T^2. For the second term, the proof is trivial, since $(n_T^{-1}\Sigma_{i\in T}Y_i)^2$ is just $\hat{\mu}_T^2$, which is a continuous function of $\hat{\mu}_T$. Consistency of $\hat{\mu}_T$ for μ_T follows from the fact that $E(\hat{\mu}_T)=\mu_T$, and $V(\hat{\mu}_T)\to 0$, as $N\to\infty$ which follows from (2). Hence, $\hat{\mu}_T^2 \xrightarrow{P} \mu_T^2$. Now, to prove consistency of $(n_T^{-1}\Sigma_{i\in T}Y_i^2)$ for (i), we need to show that $V(n_T^{-1}\Sigma_{i\in T}Y_i^2)\to 0$, as $N\to\infty$. This variance can be found by the law of variance decomposition. First, we may note that $E(n_T^{-1}\Sigma_{i\in T}Y_i^2 \mid r)=n_T^{-1}\Sigma_{i\in T}y'_{iT}$ and similarly $V(n_T^{-1}\Sigma_{i\in T}Y_i^2 \mid r)=n_T^{-2}\Sigma_{i\in T}\upsilon'_{iT}$. Now

$$V(n_T^{-1}\Sigma_{i\in T}Y_i^2)=E[V(n_T^{-1}\Sigma_{i\in T}Y_i^2 \mid r)]+V[E(n_T^{-1}\Sigma_{i\in T}Y_i^2 \mid r)]=\frac{E(\upsilon'_{iT})}{n_T}+\frac{N-n_T}{N-1}\frac{V(y'_{iT})}{n_T}=$$

$$=\frac{E(\upsilon'_{iT})}{p_T N}+\frac{V(y'_{iT})}{p_C^{-1}p_T(N-1)} \quad \text{(iv)}$$

Here, we made use of the fact that marginally with respect to treatment assignment $n_T^{-1}\Sigma_{i\in T}\upsilon'_{iT}$ is the sample mean of υ'_{iT} in a simple random sample (T) from the study population; hence its expectation is the population mean $N^{-1}\Sigma_{i=1}^{N}\upsilon'_{iT}=E(\upsilon'_{iT})$. Similarly, $n_T^{-1}\Sigma_{i\in T}y'_{iT}$ is the sample mean of y'_{iT} in T; hence, its variance is the population variance of y'_{iT} over the sample size (n_T), multiplied by the finite population correction $(N-n_T)/(N-1)$. By definition of the limiting process, the numerators in (iv) do not depend on N, while the denominators increase as $N\to\infty$. Hence, $V(n_T^{-1}\Sigma_{i\in T}Y_i^2)\to 0$, as $N\to\infty$, and $(n_T^{-1}\Sigma_{i\in T}Y_i^2) \xrightarrow{P} E(\upsilon_{iT})+N^{-1}\Sigma_{i=1}^{N}y_{iT}^2$. It follows that

$$n_T^{-1}\Sigma_{i\in T}(Y_i-\hat{\mu}_T)^2 \xrightarrow{P} E(\upsilon_{iT})+N^{-1}\Sigma_{i=1}^{N}y_{iT}^2-\mu_T^2=E(\upsilon_{iT})+V(y_{iT})$$

By a similar argument, it can be shown that $n_C^{-1}\Sigma_{i\in C}(Y_i-\hat{\mu}_C)^2 \xrightarrow{P} E(\upsilon_{iC})+V(y_{iC})$, and hence $\hat{V}(\hat{\delta})/E_a[\hat{V}(\hat{\delta})] \xrightarrow{P} 1$. This result, together with the fact that $V(\hat{\delta})/E_a[\hat{V}(\hat{\delta})]$ goes in probability to (10), implies that $V(\hat{\delta})/\hat{V}(\hat{\delta})$ also goes to (10) in probability.

Appendix 3

In this Appendix, we show that in the completely randomized design, under the regularity conditions stated in Appendix 2, $[(\hat{\delta}-\delta)/\hat{V}(\hat{\delta})^{1/2}]$ goes in distribution to a normal random variable with mean zero and variance given by (10). We will use notation $N(a,b)$ to denote a normal random variable with mean a and variance b. Also, let $d=m_T-m_C$. Our first step is to prove the Central Limit Theorem (CLT) for d that is, to show that

$$[(d-\delta)/V_a(d)^{1/2}] \xrightarrow{D} N(0,1)$$

Here "D" denotes convergence in distribution, and $V_a(d)$ is the asymptotic variance of d that can be obtained from (5) by setting $\upsilon_{iT}=\upsilon_{iC}=0,\forall i$, noting that $N-1\to N$ as $N\to\infty$, and also noting that n_T and n_C can be written as $p_T N$ and $p_C N$

$$V_a(d)=p_C\,V(y_{iT})/p_T N \;+\; p_T\,V(y_{iC})/p_C N+2Cov(y_{iT},y_{iC})/N$$

Now, to prove the CLT for d, first note that m_T is a sample mean of y_{iT} in a simple random sample (T) from a finite population (Ω); hence asymptotic normality of m_T follows from the finite-population CLT, well-known in the survey sampling literature [14]. Of course, same applies to m_C. However, asymptotic normality of d does not immediately follow because m_T and m_C are correlated. In fact, as shown in Appendix 1 $Corr(m_T,m_C)=-Corr(y_{iT},y_{iC})$. Nevertheless, if we could define for each subject in Ω a number U_i such that d could be written as a sample mean of U_i in T, that is $d=n_T^{-1}\Sigma_{i\in T}U_i$, then asymptotic normality

of d would immediately follow from the finite-population CLT applied to U_i. To find a general expression for U_i, note that the total of y_{iC} in C can be written as the total of y_{iC} in Ω, which is $N\mu_C$ minus the total of y_{iC} in T, which is $\sum_{i\in T} y_{iC}$. Hence, m_C can be written as $m_C = (N\mu_C - \sum_{i\in T} y_{iC})n_C^{-1} = n_T^{-1}\sum_{i\in T}(N\mu_C n_C^{-1} - n_T n_C^{-1} y_{iC})$, and of course by definition, $m_T = n_T^{-1}\sum_{i\in T} y_{iT}$. Hence, $d = m_T - m_C = n_T^{-1}\sum_{i\in T}[y_{iT} - (N\mu_C n_C^{-1} - n_T n_C^{-1} y_{iC})]$. It follows that a general expression for U_i exists and is given by

$$U_i = y_{iT} - (N\mu_C n_C^{-1} - n_T n_C^{-1} y_{iC}) = y_{iT} - (p_C^{-1}\mu_C - p_T p_C^{-1} y_{iC})$$

This completes the proof of the CLT for d, with the required regularity conditions for U_i satisfied by those given in Appendix 2 (see Cochran 1977, p39-40) [14]. Thus we have established that under our regularity conditions, $[(d-\delta)/V_a(d)^{1/2}] \xrightarrow{D} N(0,1)$. This also proves the CLT for $\hat{\delta}$ under the assumption of deterministic potential outcomes because this assumption implies $d = \hat{\delta}$.

If potential outcomes are stochastic and $\upsilon_{iT}^{\min}$, $\upsilon_{iC}^{\min}$ are positive (which entails practically no loss of generality since both can be arbitrarily close to zero), then given any particular treatment assignment (r), $\hat{\mu}_T$ is a mean of independent, but not necessarily identically distributed random variables, and likewise for $\hat{\mu}_C$. Given a particular treatment assignment r, $\hat{\mu}_T$ and $\hat{\mu}_C$ have expectations m_T and m_C and variances $n_T^{-2}\sum_{i\in T}\upsilon_{iT}$ and $n_C^{-2}\sum_{i\in C}\upsilon_{iC}$, respectively, and for large N, for all r approximately $[(\hat{\mu}_T - m_T)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT})^{1/2} \mid r] \sim N(0,1)$ and $[(\hat{\mu}_C - m_C)/(n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2} \mid r] \sim N(0,1)$. Here, normality follows from Lyapunov CLT [15] and our regularity conditions stated in Appendix 2. Because $\hat{\mu}_T$ and $\hat{\mu}_C$ are independent given r, we also have for large N for all r approximately

$$[(\hat{\delta}-d)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2} \mid r] \sim N(0,1)$$

Adding to both sides of the last expression $(d-\delta)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}$, which is a constant given r, we obtain

$$\left[\frac{(\hat{\delta}-\delta)}{(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}} \mid r\right] \sim N\left[\frac{(d-\delta)}{(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}}, 1\right]$$

Next we note that the marginal with respect to treatment assignment distribution of $[(\hat{\delta}-\delta)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}]$ is a mixture of $M = N!/n_T!n_C!$ conditional distributions, where M is the total number of possible treatment assignments, M^{-1} is the mixing probability, and each distribution in the mixture is normal with unit variance and mean $(d-\delta)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}$. Now we need to show that marginally with respect to treatment assignment, the distribution of these means is asymptotically normal. As we already showed, marginally with respect to treatment assignment $[(d-\delta)/V_a(d)^{1/2}] \xrightarrow{D} N(0,1)$. Also

$$[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]/[n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC}] \xrightarrow{P} 1$$

It follows from the last result and from Slutsky's Theorem [7] that the marginal asymptotic distribution of $(d-\delta)/(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}$ is the same as that of $(d-\delta)/[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]^{1/2}$. Since n_T and n_C can be written as $p_T N$ and $p_C N$, respectively, we note that $V(d)^{1/2}/[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]^{1/2}$ does not depend on N, and write

$$\frac{(d-\delta)}{[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]^{1/2}} = \frac{V_a(d)^{1/2}}{[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]^{1/2}} \frac{(d-\delta)}{V_a(d)^{1/2}} \xrightarrow{D}$$

$$\frac{V_a(d)^{1/2}}{[n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})]^{1/2}} N(0,1) = N\left[0, \frac{V_a(d)}{n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})}\right]$$

$$\frac{(d-\delta)}{(n_T^{-2}\sum_{i\in T}\upsilon_{iT} + n_C^{-2}\sum_{i\in C}\upsilon_{iC})^{1/2}} \xrightarrow{D} N\left[0, \frac{V_a(d)}{n_T^{-1}E(\upsilon_{iT}) + n_C^{-1}E(\upsilon_{iC})}\right]$$

From this result and from the fact that compounding Gaussian distributions with same variances and with means distributed according to a Gaussian distribution results in another Gaussian distribution [16] it follows that marginally with respect to treatment assignment

$$\frac{(\hat{\delta}-\delta)}{(n_T^{-2}\Sigma_{i\in T}\upsilon_{iT}+n_C^{-2}\Sigma_{i\in C}\upsilon_{iC})^{1/2}}\xrightarrow{D}N[0,1+\frac{V_a(d)}{n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})}]$$

The variance of this limiting distribution is simply the expectation of conditional variance, which is unity (since each conditional variance in the mixture is equal to unity), plus the variance of conditional expectations. Also, since marginally with respect to treatment assignment $[n_T^{-2}\Sigma_{i\in T}\upsilon_{iT}+n_C^{-2}\Sigma_{i\in C}\upsilon_{iC}]/[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]\xrightarrow{P}1$, we have marginally

$$\frac{(\hat{\delta}-\delta)}{[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]^{1/2}}\xrightarrow{D}N[0,1+\frac{V_a(d)}{n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})}]$$

Furthermore, from Appendix 2, marginally with respect to treatment assignment

$$\hat{V}(\hat{\delta})/[n_T^{-1}E(\upsilon_{iT})+n_T^{-1}V(y_{iT})+n_C^{-1}E(\upsilon_{iC})+n_C^{-1}V(y_{iC})]\xrightarrow{P}1$$

Here, $[n_T^{-1}E(\upsilon_{iT})+n_T^{-1}V(y_{iT})+n_C^{-1}E(\upsilon_{iC})+n_C^{-1}V(y_{iC})]=E_a[\hat{V}(\hat{\delta})]$ is the asymptotic form of (8). By Slutsky's Theorem, the marginal asymptotic distribution of $(\hat{\delta}-\delta)/\hat{V}(\hat{\delta})^{1/2}$ is the same as that of $(\hat{\delta}-\delta)/\{E_a[\hat{V}(\hat{\delta})]\}^{1/2}$. Noting that $[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]^{1/2}/\{E_a[\hat{V}(\hat{\delta})]\}^{1/2}$ does not depend on N, we can write that marginally with respect to treatment assignment

$$\frac{(\hat{\delta}-\delta)}{\{E_a[\hat{V}(\hat{\delta})]\}^{1/2}}=\frac{[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]^{1/2}}{\{E_a[\hat{V}(\hat{\delta})]\}^{1/2}}\times\frac{(\hat{\delta}-\delta)}{[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]^{1/2}}\xrightarrow{D}$$

$$\frac{[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})]^{1/2}}{\{E_a[\hat{V}(\hat{\delta})]\}^{1/2}}N[0,1+\frac{V_a(d)}{n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})}]=N\{0,\frac{n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})+V_a(d)}{E_a[\hat{V}(\hat{\delta})]}\}$$

$$\frac{(\hat{\delta}-\delta)}{\hat{V}(\hat{\delta})^{1/2}}\xrightarrow{D}N\{0,\frac{n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})+V_a(d)}{E_a[\hat{V}(\hat{\delta})]}\}$$

Here, $[n_T^{-1}E(\upsilon_{iT})+n_C^{-1}E(\upsilon_{iC})+V_a(d)]/E_a[\hat{V}(\hat{\delta})]$ is given by (10) and hence cannot exceed unity.