




Article

# An Epidemiologic Approach for Estimating Risk Reduction and Asymptotic Power on the Log-Difference Scale

Jimmy T. Efird <sup>1,2</sup> 

- <sup>1</sup> Cooperative Studies Program Coordinating Center, VA Boston Healthcare System, Lafayette City Center, 2 Avenue de Lafayette, Boston, MA 02111, USA; jimmy.efird@va.gov or jte38@case.edu; Tel.: +1-650-248-8282  
<sup>2</sup> Department of Radiation Oncology, School of Medicine, Case Western Reserve University, Cleveland, OH 44206, USA

**Abstract:** When comparing the efficacy or harmfulness of two groups (e.g., drugs, devices, assays, interventions, environmental toxins), it is important to minimize bias by making this comparison with respect to a common referent-control group, assuming random allocation. Under such a scenario, one can estimate risk reduction for a new therapy on a log-difference, relative effect scale. The current manuscript reviews the large-sample framework for this conditionally independent comparison and demonstrates how to estimate test power for a given sample size.

**Keywords:** clinical trials; common referent-control; conditional independence; multiplicity adjustment; power; risk reduction; sample size

## 1. Introduction

A frequently used clinical trial design is to directly compare a new treatment against a standard therapeutic compound. However, many years may have passed since the standard compound was originally approved. During this time, variations in the disease process, manufacturing changes, or practice deviations may have occurred, such that the standard treatment is no longer efficacious, or is effective to a much lesser degree than originally marketed.

Observing a statistically significant absolute effect (risk) reduction (i.e., the standard minus new therapy event rate) may not necessarily prove the clinical usefulness of the new treatment unless the clinical trial design includes a common comparison group as a referent-control. This contrast allows for the side-by-side evaluation of the new and standard agents in the context of relative effect reduction on the log-difference scale. Importantly, the use of a shared referent-control arm minimizes bias and threats to internal validity compared with assessing the absolute risk difference.

Power and sample size are important components in the planning of a clinical trial. However, the literature is largely silent on the topic of a conditionally independent shared comparison arm when assessing risk reduction on the log-difference, relative effect scale.

## 2. Preliminaries

### 2.1. Conditionally Independent, Large-Sample Distribution

Consider a test to compare two relative effect estimates (REEs), denoted as  $(\hat{\theta}_1, \hat{\theta}_2)$ , on the log-difference scale, with respect to a common referent-control group. Approximate log-normally distributed REEs include relative risks (RR), odds ratios (OR), and hazard ratios (HR). A variance-stabilizing, logarithmic transformation tends to make the difference



Academic Editor: Kenneth C. Land

Received: 23 February 2025

Revised: 6 April 2025

Accepted: 29 April 2025

Published: 1 May 2025

**Citation:** Efird, J.T. An Epidemiologic Approach for Estimating Risk Reduction and Asymptotic Power on the Log-Difference Scale. *Int. J. Environ. Res. Public Health* **2025**, *22*, 719. <https://doi.org/10.3390/ijerph22050719>

**Copyright:** © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of REEs closer to a normal distribution by reducing any potential relationship between the variance and mean. The application of this transformation is especially useful when dealing with probabilities near 0 or 1.

Assuming that (a) the standard error (SE) for the logarithm of  $(\hat{\vartheta}_1)$  does not differ considerably from that for  $(\hat{\vartheta}_2)$ , (b) corresponding sample sizes  $(n_{\hat{\vartheta}_1}, n_{\hat{\vartheta}_2})$  are reasonably large, (c)  $\hat{\vartheta}_1$  and  $\hat{\vartheta}_2$  with a common referent-control arm are asymptotically consistent, but not necessarily unbiased, estimates of the true parameters, and (d) the large-sample distribution for the logarithm of  $(\hat{\vartheta}_1, \hat{\vartheta}_2)$  is approximately Gaussian, with estimated event rates for both groups not being too close to 0 or 1, it follows from the Le Cam–Lévy–Ferguson “Conditional Martingale Limit Theorem” that we obtain the following equation:

$$\lim_{(n_{\hat{\vartheta}_1}, n_{\hat{\vartheta}_2}) \rightarrow \infty} \frac{\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right) - E\left[\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right)\right]}{\sqrt{\text{Var}\left[\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right)\right]}} \sim N(0, 1), \quad (1)$$

where  $E[*]$  represents the expectation, and  $N(0, 1)$  is the standard normal distribution [1]. The single tilde, “ $\sim$ ”, indicates asymptotic equivalence of the same magnitude, under the assumption that the underlying Martingale processes are stationary, ergodic, and F-measurable [2], with finite variance.

The cumulative distribution function for  $\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right)$  divided by that for a standard normal curve approaches unity for large sample sizes. Under asymptotic regularity conditions,  $\hat{\vartheta}_1$  is conditionally independent of  $\hat{\vartheta}_2$ , given a common referent-control group. That is, the probability (A) for an observation in the numerator of  $\hat{\vartheta}_1$  is not affected by the probability (B) for an observation in the numerator of  $\hat{\vartheta}_2$ , given the probability (C) for a shared observation in the denominators of  $\hat{\vartheta}_1$  and  $\hat{\vartheta}_2$ . Mathematically stated, it follows that  $(A \perp\!\!\!\perp B \mid C)$  is equivalent to  $(B \perp\!\!\!\perp A \mid C)$ .

The denominator of the limit described above, denoted as  $(\zeta)$ , yields the square root of the sample variance estimate of  $[\log(\hat{\vartheta}_1) - \log(\hat{\vartheta}_2)]$ , which may be approximated as follows [3]:

$$\zeta \approx \sqrt{\{SE[\log(\hat{\vartheta}_1)]\}^2 + \{SE[\log(\hat{\vartheta}_2)]\}^2}. \quad (2)$$

## 2.2. Confidence Interval Method for Estimating the Sample Variance

Lower and upper  $\alpha$ -level confidence intervals (CIs) for the indicated REEs are given as  $LCI[*]$  and  $UCI[*]$ . That is, for  $(i = 1, 2)$ , we obtain the following expression:

$$CI_{\alpha}(\vartheta_i) = e^{\log(\hat{\vartheta}_i) \pm z_{(\alpha/2)} SE[\log(\hat{\vartheta}_i)]}, \quad (3)$$

where  $(\alpha)$  refers to the probability of rejecting the null hypothesis when it is true, and  $z_{(\alpha/2)}$  corresponds to the  $100(1 - \alpha/2)$  percentile of  $N(0, 1)$  [e.g.,  $z_{(0.05/2)} = 1.96$ ]. Using the above bounds and rearranging, it readily follows that we obtain the following equation:

$$\zeta \approx \sqrt{\left\{ \frac{\log[UCI(\hat{\vartheta}_1)] - \log[LCI(\hat{\vartheta}_1)]}{2z_{(\frac{\alpha}{2})}} \right\}^2 + \left\{ \frac{\log[UCI(\hat{\vartheta}_2)] - \log[LCI(\hat{\vartheta}_2)]}{2z_{(\frac{\alpha}{2})}} \right\}^2}. \quad (4)$$

### 2.3. Alternative Computational Formula for $SE[\log(\widehat{RR})]$

Let  $(a_{i,j})$  denote the  $(i^{th}, j^{th})$  cell sizes for a contingency table with corresponding probabilities of  $(\pi_{i,j})$ . Here, the row exposures and column outcomes are indexed by  $(i)$  and  $(j)$ . The relative risk estimate  $(\widehat{RR})$ , also known as the risk ratio, is defined as follows:

$$\widehat{RR} = \frac{a_{11}/a_{1+}}{a_{21}/a_{2+}}, \quad (5)$$

where  $(a_{1,1})$  and  $(a_{2,1})$  follow a binomial distribution. By applying the delta method (Taylor series approximation for large sample sizes), we obtain the following equations:

$$Var\left\{\log(\widehat{RR})\right\} = Var\left\{\log\left(\frac{a_{11}}{a_{1+}}\right)\right\} + Var\left\{\log\left(\frac{a_{21}}{a_{2+}}\right)\right\} \quad (6)$$

$$\approx \left(\frac{a_{1+}}{a_{11}}\right)^2 \left\{\frac{\frac{a_{11}}{a_{1+}}\left(1 - \frac{a_{11}}{a_{1+}}\right)}{a_{1+}}\right\} + \left(\frac{a_{2+}}{a_{21}}\right)^2 \left\{\frac{\frac{a_{21}}{a_{2+}}\left(1 - \frac{a_{21}}{a_{2+}}\right)}{a_{2+}}\right\} \quad (7)$$

$$\approx \frac{1}{a_{11}} - \frac{1}{a_{1+}} + \frac{1}{a_{21}} - \frac{1}{a_{2+}}. \quad (8)$$

Accordingly, we obtain the following equation:

$$SE[\log(\widehat{RR})] \approx \sqrt{\frac{1}{a_{11}} - \frac{1}{a_{1+}} + \frac{1}{a_{21}} - \frac{1}{a_{2+}}}.$$

The advantage of this direct formula is the increased computational speed gained by foregoing the CI method for estimating  $SE[\log(\widehat{RR})]$ . In the case of ORs and HRs, one can derive similar computational approximations using the delta method.

### 2.4. Percentage Relative Effect (Risk) Reduction

When  $(\hat{\vartheta}_1 > \hat{\vartheta}_2)$ , the percentage relative effect reduction (%RLD) is computed as  $100\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2} - 1\right)$ , and vice versa when  $(\hat{\vartheta}_2 > \hat{\vartheta}_1)$ , i.e.,  $100\left(\frac{\hat{\vartheta}_2}{\hat{\vartheta}_1} - 1\right)$ . If  $(\hat{\vartheta}_1 = \hat{\vartheta}_2)$ , a null result is obtained for the (%RLD). For example, if  $(\hat{\vartheta}_1 = 0.80)$  and  $(\hat{\vartheta}_2 = 0.60)$ , then  $(\%RLD) = 100\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2} - 1\right) = 100\left(\frac{0.80}{0.60} - 1\right) = 100(1.33 - 1) = 33$ .

### 2.5. Null Hypothesis and $p$ -Value

In the case of a null hypothesis ( $H_0$ ),  $\log(\hat{\vartheta}_1)$  and  $\log(\hat{\vartheta}_2)$  are deemed to be equal, versus the alternative scenario in which they differ [3]. That is,  $H_0: \log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right) = 0$ , which is equivalent to a null %RLD, compared with  $H_1: \log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right) \neq 0$ . In the case of ( $H_0$ ), both  $\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right)$  and  $E\left[\log\left(\frac{\hat{\vartheta}_1}{\hat{\vartheta}_2}\right)\right]$  equal zero.

A  $p$ -value for the log-difference of two REEs is estimated as follows:

$$P \sim 2 \cdot \{1 - \Phi(z)\}, \quad (9)$$

where

$$\Phi(z) = \int_{-\infty}^z \frac{e^{-x^2/2}}{\sqrt{2 \cdot \pi}} dx \quad (10)$$

and

$$z \sim \frac{\log(\hat{\vartheta}_1) - \log(\hat{\vartheta}_2)}{\zeta}. \quad (11)$$

### 2.6. Confidence Interval Decision Criteria

The disjoint CI percentage corresponding to an equivalent  $\alpha$ -level test is computed as follows:

$$\left[ 1 - 2\Phi\left(-Z_{(\alpha/2)} \frac{\sqrt{2}}{2}\right) \right] \times 100\%. \quad (12)$$

This yields slightly wider regions to compensate for the pooled (versus non-pooled) sample variance [4]. For example, given that  $\alpha = 0.01, 0.05$ , and  $0.10$ , the difference between the estimates  $\log(\hat{\theta}_1)$  and  $\log(\hat{\theta}_2)$  is declared as statistically significant if the 93.1452%, 83.4224%, and 75.5206% CIs for  $(\hat{\theta}_1, \hat{\theta}_2)$  do not overlap.

Analogously, the difference between the estimates  $\log(\hat{\theta}_1)$  and  $\log(\hat{\theta}_2)$  may be declared as statistically significant when the following equation excludes unity:

$$CI_{\alpha}\left(\frac{\theta_1}{\theta_2}\right) = e^{\log(\frac{\theta_1}{\theta_2}) \pm \{z_{(\alpha/2)}\}\zeta} \quad (13)$$

## 3. Power and Sample Size

### 3.1. Power

The power of a study indicates how frequently a statistical test will detect the falsehood of an underlying null hypothesis when it is false (i.e.,  $\Pr(\text{reject } H_0 | H_0 \text{ false})$ ). Let  $(\beta)$  denote the probability of failing to reject the null the hypothesis when it is false (i.e.,  $\Pr(\text{accept } H_0 | H_0 \text{ false})$ ). It follows that we obtain the following equation:

$$z_{(\beta)} \sim \left\{ \left[ \frac{\log(RLD + 1)}{\zeta} \right] - z_{(\alpha/2)} \right\}. \quad (14)$$

The area to the left of  $z_{(\beta)}$  under a standard normal distribution yields the desired power at the  $\alpha$ -level of statistical significance for a two-sided test. When use of a one-tailed test is desired,  $z_{(\alpha)}$ , rather than  $z_{(\alpha/2)}$ , is used in this formula.

### 3.2. Sample Size and Variability

The confidence interval method may be used to approximate the standard error for the logarithm of the respective relative effect estimates  $(\hat{\theta}_1, \hat{\theta}_2)$  corresponding to an initial rxc (row by column) = (n) contingency table, where (n) denotes the total sample size (i.e., combined rows). Cell frequencies of the pilot sample table are increased multiplicatively (in an iterative fashion) to obtain the corresponding standard errors for different sample sizes and corresponding power.

Using this method, it is important that the initial  $3 \times 2$  contingency table be sufficiently large to accurately estimate the sample variability within a specified fraction of the population variance [5]. As a rule of thumb, the proportional width of the  $\alpha$ -level CI for the population variance of a continuous variable is defined as follows:

$$(n-1) \left( \frac{1}{\chi^2_{(1-\alpha/2), (n-1)}} - \frac{1}{\chi^2_{(\alpha/2), (n-1)}} \right), \quad (15)$$

This value should be as large as the sample variance, where  $\chi^2_{(\alpha/2), (n-1)}$  is the  $100(1 - \alpha/2)$  percentile of the chi-squared distribution with  $(n - 1)$  degrees of freedom. When  $\alpha = 0.05$ , this corresponds to a minimum sample size of 38. However, in practice, a more conservative sample size of 90 (30 per treatment row) is typically chosen as the starting point to account for the binary outcomes.

### 3.3. Simulation of Observations from a Multinomial Distribution

Given (n) distinct trials, the probability that a mutually exclusive set of (k) non-negative random variates ( $A_1, A_2, \dots, A_k$ ) takes on a particular value ( $a_1, a_2, \dots, a_k$ ) is given as follows:

$$P_n(A_1 = a_1, A_2 = a_2, \dots, A_k = a_k) = \frac{n!}{\prod_{i=1}^k a_i!} \prod_{i=1}^k \pi_i^{a_i}, \quad (16)$$

where

$$P_n(A_i = a_i) = \pi_i, \sum_{i=1}^k \pi_i = 1, \text{ and } \sum_{i=1}^k a_i = n. \quad (17)$$

This probability is known as a multinomial distribution because of the following expression:

$$\frac{n!}{\prod_{i=1}^k a_i!} \prod_{i=1}^k \pi_i^{a_i} = (a_1 + a_2 + \dots + a_k)^n, \quad (18)$$

where the latter is a multinomial series.

The multinomial distribution, as illustrated in Example 2, plays an important role in validating the asymptotic normality of the conditional test statistic.

## 4. Computational Methods

Analyses were performed and validated in SAS 9.4 (Cary, NC).

## 5. Example 1—Comparison of Potassium and Sodium Salts

Potassium salt substitutes can be helpful for lowering sodium intake and controlling high blood pressure. Among a cohort of borderline hypertensive (but otherwise healthy) patients, a team of clinical epidemiologists aimed to determine if randomization to a potassium salt substitute (1500 mg/d) reduces the 24-month risk of major adverse cardiovascular outcomes and non-cancer death compared with a formulary of sodium salt (1500 mg/d), with a common referent-control arm of 2300 mg/d sodium salt.

In a 1:1:1 pilot clinical trial, a 9.52% relative effect (risk) reduction on the log-difference scale was observed at 24 months post-randomization (Table 1). Based on this promising result, the team decided to conduct a larger population-based phase III clinical trial. To achieve at least 80% power for a similar %RLD (discounting dropouts), the new pivotal trial would need to randomize 36,835 patients per study arm at the  $\alpha = 0.05$  level of statistical significance (Table 2; see Appendix A for SAS code).

**Table 1.** Pilot analysis (N = 500 participants per arm).

Treatment Regimen (T)		MACE Outcome at 24 Months Post-Baseline n (%)		Results <sup>†</sup>
		Yes	No	
$T_0$	2300 mg/d <sup>§</sup> Sodium Salt	50 (10)	450 (90)	$\hat{\theta}_1(T_0 : T_2) = 1.1905$ ( $CI_{0.05} = 0.8052 - 1.7602$ ) $\hat{\theta}_2(T_0 : T_1) = 1.0870$ ( $CI_{0.05} = 0.7428 - 1.5906$ ) $\frac{\hat{\theta}_1}{\hat{\theta}_2} = 1.0952$ ( $CI_{0.05} = 0.6346 - 1.8903$ ) %RLD = 9.52%; $P = 0.74395$
$T_1$	1500 mg/d Sodium Salt	46 (9)	454 (91)	
$T_2$	1500 mg/d Potassium Salt	42 (8)	458 (92)	

<sup>†</sup> Estimates are depicted as rounded values. <sup>§</sup> Common referent-control. Baseline = initiation of treatment regimen.  $CI_{0.05}$  = 95% confidence interval. MACE = major cardiovascular event. mg/d = milligrams per day. n = cell frequency. P = p-value for risk reduction.  $\hat{\theta}_i$  = risk ratio for group (i). %RLD = percentage relative effect (risk) reduction on the log-difference scale.

**Table 2.** Power analysis for a planned population-based phase III clinical trial.

Sample Size Per Arm	Comparison	$CI_{0.05}(\hat{\theta})^{\dagger}$	$SE\left[\log(\hat{\theta})\right]^{\dagger}$	Minimal Detectable RLD (%)	Power $^{\dagger,\S}$ (%)
10,000	$T_0 : T_2$	1.0908–1.2993	0.044615	14	55.7
				12	44.4
	$T_0 : T_1$	.9982–1.1836	0.043439	10	33.4
				8	23.5
20,000	$T_0 : T_2$	1.1191–1.2664	0.031547	14	84.5
				12	73.0
	$T_0 : T_1$	1.0234–1.1544	0.030716	10	58.1
				8	41.6
30,000	$T_0 : T_2$	1.1319–1.2521	0.025758	14	95.4
				12	88.3
	$T_0 : T_1$	1.0348–1.1417	0.025080	10	75.5
				8	57.2
40,000	$T_0 : T_2$	1.1395–1.2437	0.022307	14	98.8
				12	95.4
	$T_0 : T_1$	1.0417–1.1342	0.021720	10	86.5
				8	69.6

$^{\dagger}$  Estimates depicted as rounded values.  $^{\S}$  Power for a 2-sided,  $\alpha = 0.05$  level test.  $CI_{0.05} = 95\%$  confidence interval.  $RLD$  = Relative effect (risk) reduction on the log-difference scale.  $\hat{\theta}$  = Risk ratio.  $SE$  = Standard error.

## 6. Example 2—Monte Carlo Simulation of the Conditional Martingale Limit Theorem

While a theoretical basis exists for the asymptotic normality and conditional independence of the test statistic [1,2], Monte Carlo methods can be used to validate the large-sample properties of this method.

A total of 1,000,000 trials, repeated for each sample of 900,000 patients (300,000 per arm), were drawn from a multinomial distribution using the cell probabilities given in Example 1 (with  $T_0$  denoting the common referent-control group). A standardized value was computed for each simulated observation for  $\log\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right)$  by subtracting  $E\left(\log\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right)\right)$  from this estimate and dividing it by  $\sqrt{\text{Var}\left[\log\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right)\right]}$ . Plotting these values as a histogram yields Figure 1. The corresponding statistics (mean = 0, standard deviation = 1, and approximate zero values for skewness and kurtosis) indicate that the z-scores follow a standard normal distribution, as visually confirmed by the Normal (0,1) curve overlay. Furthermore, normality was not rejected by the Kolmogorov–Smirnov, Cramer–von Mises, or Anderson–Darling tests (see Appendix B for SAS code).

For large samples, this example supports the use of normal-theory methods to estimate power for a 1:1:1 randomized design with a common referent-control arm. In effect, the covariance of  $\log(\hat{\theta}_1)$  and  $\log(\hat{\theta}_2)$  can be disregarded as the sample size increases toward infinity, allowing one to assume an independent, identically distributed (i.i.d.) standard normal test statistic.

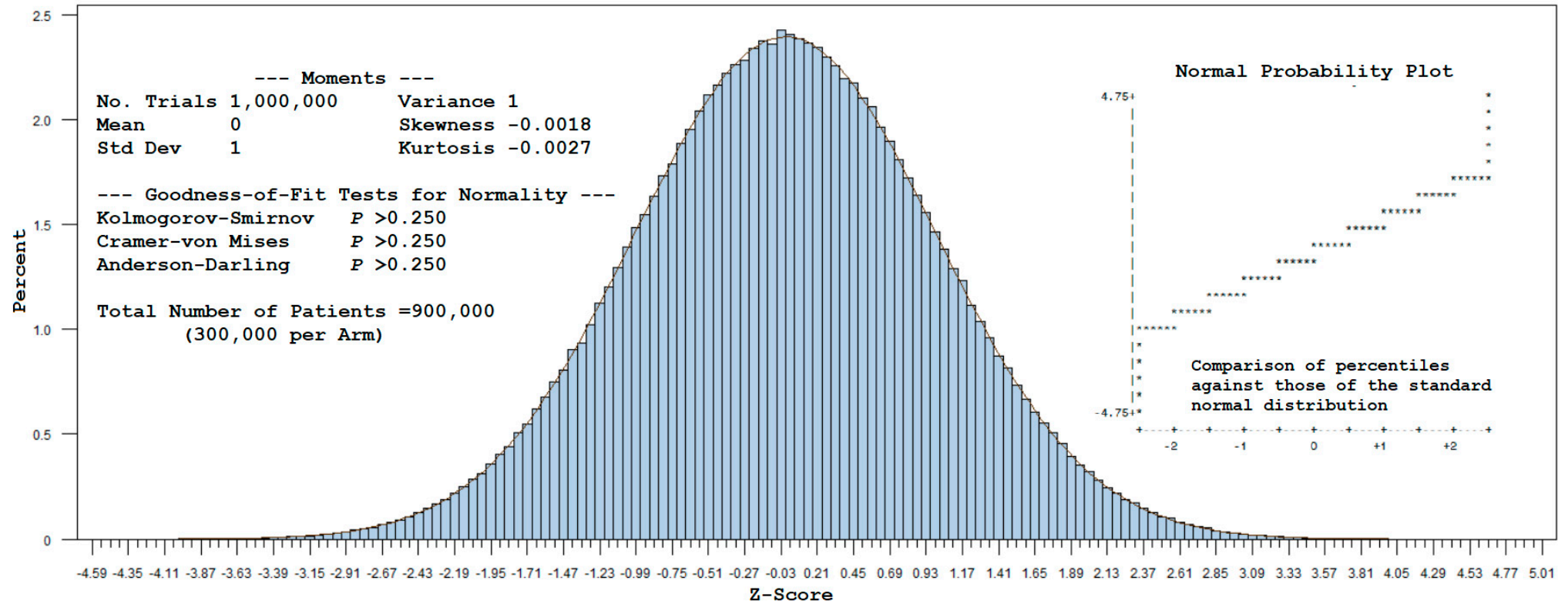


Figure 1. Histogram with a standard normal overlay for simulated z-scores, conditional on a common reference-control arm.



## 7. Discussion

### 7.1. Overview

The comparison of two drugs with respect to a common comparator (referent-control) group is a powerful and efficient approach for conducting a clinical trial. Importantly, this method preserves the randomization of the originally assigned patient groups [6]. Additionally, it allows for determining risk reduction on the log-difference scale. The current manuscript presents a novel, computer-based technique for estimating the power of the latter by multiplicatively increasing the cell sizes of an initial pilot sample. This yields the inflated variances needed to compute power for larger sample sizes while still allowing the use of traditional normal-theory methods. As reasoned below, the common referent-control design does not require multiplicity adjustment.

### 7.2. Multiplicity Adjustment

The family-wise type I error rate, which is the probability of at least one false positive test among all hypotheses being tested, is not increased with respect to a common referent-control group [7]. The associated false positive comparison-wise error rate at the individual level is also not inflated, as the “expected proportion of incorrectly rejected hypotheses will not exceed the significance level used in the individual tests” [7,8].

While a treatment is the “family” over which one typically judges the need for multiplicity, there is indistinctness regarding how this entity is defined. Multiplicity control is not required if the underlying therapies have distinct mechanisms of action [9]. That is, “if each decision to reject each individual null hypothesis depends on no more than one significance test, then none of the individual tests constitute a family”, such that an  $\alpha$ -level adjustment for any single hypothesis is unnecessary (except if one performs “disjunction testing of a joint intersection null hypothesis”) [10]. While cases exist wherein “the decision on the omnibus null will seem to contradict decisions on individual nulls”, it is important to remember that “rejecting or not rejecting a null is not certain proof that the null is false or true” [10]. Therefore, multiplicity correction for CIs is needed when the omnibus null can only be evaluated through “a set of surrogate nulls” and cannot be “generalized to testing any diversity of unconnected nulls over the course of a study” [11].

Adjustment is counterintuitive in the case of conjunctive (intersection–union) testing, in which all tests must be significant for the joint null hypothesis to be rejected [9]. However, multiplicity adjustment for CIs is advisable if the comparative therapies are related, as is the case when evaluating “different dosages or regimens of a treatment compared with the same control arm” [12]. This includes trials with collective conditions “if there are no less than two primary hypotheses, unless one assumes that there is an explicit hierarchy in the multiple hypotheses” [12].

Multiplicity adjustment of CIs in the denominator of a relative-effects (log-difference) test of two compounds is generally not performed in practice, as this is equivalent to a stand-alone interaction test with a pooled variance structure [3]. If requested by a regulatory agency, the intervals can be easily adjusted for multiplicity using the Hochberg step-up procedure [13]. However, when “such adjustments are applied unnecessarily, potentially effective treatments may be discarded prematurely” [9].

### 7.3. Limitations

A limitation of REEs is that they are ratios and tend to have a slightly skewed log-normal distribution, which creates issues when the parameters of a test statistic are estimated from the sample. According to Jensen’s inequality, the expectation will be less than or equal to its true value. To some degree, this bias is an advantage because it helps to offset the Cauchy infinite variance problem that may result from taking the ratio of two normally



distributed variables [1]. Barring extreme degenerate examples, it is noted that  $\log\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right)$  asymptotically assumes a moderately well-behaved, normal shape in the limit.

The literature offers little consistent advice regarding the lower limit for the sample size needed to satisfy the underlying large-sample assumptions of the test statistic. It is also unclear what level of inaccuracy is incurred when estimating a z-score test statistic when the sample size is small and the observations are not fully independent. By treating the logarithmic transformation of  $\frac{\hat{\theta}_1}{\hat{\theta}_2}$  as a mean difference and applying Gauss' hypergeometric series, one may heuristically conclude that the small-sample test statistic follows a pseudo Behrens–Fisher distribution with  $(v)$  degrees of freedom [14], as follows:

$$v = \frac{\left[ \frac{SD[\log(\hat{\theta}_1)]}{n_{\hat{\theta}_1}} + \frac{SD[\log(\hat{\theta}_2)]}{n_{\hat{\theta}_2}} \right]^2}{\left[ \frac{\{SD[\log(\hat{\theta}_1)]\}^2}{n_{\hat{\theta}_1} + 1} + \frac{\{SD[\log(\hat{\theta}_2)]\}^2}{n_{\hat{\theta}_2} + 1} \right]} - 2. \quad (19)$$

An alternative, albeit lower-power, approach is the Lepage test, which treats the problem as a nonparametric test for central tendency and dispersion [15]. In certain situations, nonparametric methods are better suited (less sensitive) for handling violations of independence.

Addressing the exact convergence rates for a conditional Martingale-type central limit theorem, which are ostensibly slower than  $(1/\sqrt{n})$ , is a complex topic beyond the scope of the current manuscript and is best deferred to future research [16]. Nonetheless, it is common practice in many applied scenarios to accept that the underlying data are “close enough” to normal to move forward with normal-theory methods.

Imprecise rounding is also a concern. As REEs are expressed on a logarithmic scale, with significant departures from linearity occurring near unity, it is best to avoid intermediate rounding when computing the test statistic. Fortunately, this rarely poses a problem when the operations are performed by a computer rather than by manual means.

In theory, power for a given sample size may be estimated by simulating datasets from a conditional multinomial distribution, with prescribed cell probabilities being reflective of the underlying hypothesized parameters. Power is then defined as the proportion of the simulated test statistics that fall within the “a priori” alpha rejection region. However, the simulation method is computationally intensive and introduces uncertainty in the values at the distribution tails. As large sample sizes are often needed to detect small effect sizes for risk reduction, the process may be resource-prohibitive or prone to computational errors, in which case normal-theory methods may be the better choice for computing power. While robust estimation methods are available in most standard statistical packages to simulate the covariance of REEs and corresponding test statistics involving correlated data, boundary issues regularly result in algorithmic non-convergence. The simulation of power also is problematic for heavy-tailed, Cauchy-like distributions, which may require sample sizes on the order of millions to properly characterize outlying values. Furthermore, multinomial-type power computations are predicated on the observed pilot sample effect size. Thus, simulation cannot be used to determine power for a range of minimally detectable  $RLD(\%)$  values based on a single effect size estimate. In such cases, separate pilot samples will be needed to determine the desired range of values.

#### 7.4. Future Directions

Future investigations are directed at estimating power when sample sizes are small and the assumption of conditional independence is questionable. This will include cases where it is difficult to establish the consistency, stability, and normality of underlying

data estimates and parameters. Innovative research efforts focusing on nonparametric, two-sample conditional tests and Bayesian methods also may be useful [17,18].

## 8. Conclusions

The use of a common reference-control arm is an important tool for minimizing bias in a randomized clinical trial. Assuming conditional independence, a large-sample method is presented to estimate power for risk reduction on the log-difference scale. This computationally straightforward approach offsets the uncertainty and potential bias associated with tests of absolute risk reduction.

**Funding:** The author reports that there is no external funding associated with the work featured in this article. The opinions presented in this manuscript do not necessarily represent those of the VA or the United States Federal Government. All examples provided herein are hypothetical and not intended to reflect any real individuals, communities, or non-profit/corporate/governmental entities.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The insightful comments from Genevieve Dupuis, Kaitlin Cassidy, and Janet M. Grubber during the writing and revision of this manuscript are greatly appreciated. SAS programming was validated by Maria Androsenko.

**Conflicts of Interest:** The author declares no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ASCVD	Atherosclerotic cardiovascular disease
CI	Confidence interval
LCI	Lower confidence interval
Log	Logarithm
Mg/day	Milligrams per day
$P$	$p$ -value
REE	Relative effect estimate
RLD	Relative effect reduction on the log-difference scale
SD	Standard deviation
SE	Standard error
UCI	Upper confidence interval

## Appendix A. SAS Code for Determining Power and Sample Size, as Shown in Table 2

```

options ls=180 ps=40 cleanup compress=yes reuse=yes nocenter nodate mprint;
%global j k ss cnt;
/** Input 3x2 pilot data
***/;
data a;
  do a=1 to 3;
    do b=1 to 2;
      input wt @@;
      output;
    end;
  end;
cards;
50 450
46 454
42 458
;
/** Determine pilot sample count ***/;
data count;
  set a;
  if _n_ <= 2;
  retain cnt 0;
  cnt = cnt + wt;
  if _n_ = 2;
  call symput("cnt", cnt);
%put value = &cnt.;
data master(drop=x);
  x = -1;
/** Begin macro loop to estimate power ***/;
%macro rrr(rr, star1, star2);
  ods output "Relative Risk Estimates" = &rr;
  proc freq data=a2;
    weight wt;
    tables a*b/all;
    &star1 where a in (1,3);
    &star2 where a in (1,2);
  run;
  ods output close;
%mend rrr;
%macro loopxy(start1=, end1=, by1=, start2=, end2=, by2=);
  %local x;
  %do x=1 %to %eval(%sysfunc(ceil(%sysvalf((&end1-&start1)/&by1)))+1) ;
    %let valuex = %sysvalf((&start1-&by1)+(&by1*&x)) ;
    %if &valuex <= &end1 %then %do;

```

```

                                %put &valuex;
                                %end;

%local y;
%do y=1 %to %eval(%sysfunc(ceil(%sysevalf((&end2-&start2)/&by2)))+1) ;
    %let valuey=%sysevalf((&start2-&by2)+(&by2*&y)) ;
    %if &valuey<=&end2 %then %do;
        %put &valuey;
    %end;

%let ss=&valuex;
data a2;
    set a;
    wt=&valuex*wt;
%rrr(%str(rr1),%str(),%str(*));
%rrr(%str(rr2),%str(*),%str());
data rrr;
    merge rr1(rename=(value=rr1 lowercl=lci1 uppercl=uci1) keep=value lowercl uppercl)
          rr2(rename=(value=rr2 lowercl=lci2 uppercl=uci2) keep=value lowercl uppercl);
    if _n_=2;
    zalphad2=probit(1-0.05/2);
    pctrld=&valuey;
    rrr=(pctrld/100)+1;
    lrrr=log(rrr);
    se1=((log(uci1)-log(lci1))/(2*zalphad2));
    se2=((log(uci2)-log(lci2))/(2*zalphad2));
    zbeta=(lrrr/sqrt(se1**2+se2**2))-zalphad2;
    power=probnorm(zbeta);
    ss=&ss*&cnt;
    /** Combine results ***/;
data master;
    set master rrr;
    if rrr>0;

run;

%end;

%end;

%mend loopxy;
%loopxy(start1=20,end1=80,by1=20,start2=8.0,end2=14,by2=2);
/** The first loop controls the multiplication factor for determining sample size ***/;
/** The second loop controls the minimal detectable RLD percentage ***/;
proc sort data=master;
    by ss descending pctrld;
proc print data=master;
run;

```

## Appendix B. SAS Code for Simulating the Conditional Central Limit Theorem (Example 2)

```

options ls=180 ps=40 cleanup nocenter compress=yes reuse=yes nodate nonumber
        mprint bufsize=12288;
/** Set random seed, specify cell probabilities, and simulated multinomial observations ***/;
proc iml symsize=1048576 worksize=1048576;
    call randseed(714);
    c={"a11","a12","a21","a22","a31","a32"};
    prob={ .033333333333 .300000000000
           .030666666667 .302666666667
           .028000000000 .305333333333};
    x=randmultinomial(1000000,900000,prob);
    create rr_data from x[colname={ "a11","a12",
                                   "a21","a22",
                                   "a31","a32"}];

    append from x;
    close rr_data;
    free c prob x;
quit;
/** Compute logarithm for ratio of RR1 and RR2, assuming common referent-control arm ***/;
data a(drop=a11 a12 a21 a22 a31 a32);
    set rr_data;
    zalphad2=probit(1-0.05/2);
    theta1top=a11/(a11+a12);
    theta1bot=a31/(a31+a32);
    rrtheta1=theta1top/theta1bot;
    theta2top=a11/(a11+a12);
    theta2bot=a21/(a21+a22);
    rrtheta2=theta2top/theta2bot;
    lrrr=log(rrtheta1/rrtheta2);
proc univariate data=a noprint;
    var lrrr;
    output out=out1 mean=mean std=sd;
/** Compute standardized zscore ***/;
data a;
    if _n_=1 then set out1;
    else          set a;
    zscore=(lrrr-mean)/sd;
    if _n_>1;
/** Plot and compute tests of normality, with standard normal curve overlay ***/;
proc univariate data=a plots;
    histogram zscore / normal(mu=0 sigma=1);
run;

```

## References

1. Le Cam, L. *Asymptotic Methods in Statistical Decision Theory*; Springer: New York, NY, USA, 1986.
2. Peligrad, M. Conditional central limit theorem via Martingale approximation. *arXiv* **2011**. [[CrossRef](#)]
3. Altman, D.; Bland, J. Interaction revisited: The difference between two estimates. *Br. Med. J.* **2003**, *326*, 219. [[CrossRef](#)] [[PubMed](#)]
4. Knol, M.; Pestmen, W.; Grobbee, D. The (mis)use of overlap of confidence intervals to assess effect modification. *Eur. J. Epidemiol.* **2011**, *26*, 253–254. [[CrossRef](#)] [[PubMed](#)]
5. Teare, M.; Dimairo, M.; Shephard, N.; Hayman, A.; Whitehead, A.; Walters, J. Sample size requirements to estimate key design parameters from external pilot randomized controlled trials: A simulation study. *Trials* **2014**, *15*, 264–277. [[CrossRef](#)] [[PubMed](#)]
6. Kim, H.; Gurrin, L.; Ademi, Z.; Liew, D. Overview of methods for comparing the efficacies of drugs in absence of head-to-head clinical trial data. *B.J.C.P* **2013**, *77*, 116–121. [[CrossRef](#)] [[PubMed](#)]
7. Parker, R.; Weir, C. Non-adjustment for multiple testing in multi-arm trials of distinct treatments: Rationale and justification. *Clin. Trials* **2020**, *17*, 562–566. [[CrossRef](#)] [[PubMed](#)]
8. Bender, R.; Lange, S. Adjusting for multiple testing—When and how? *J. Clin. Epidemiol.* **2001**, *54*, 343–349. [[CrossRef](#)] [[PubMed](#)]
9. Molloy, S.; White, I.; Nunn, A.; Hayes, R.; Wang, D.; Harrison, S. Multiplicity adjustments in parallel-group multi-arm trials sharing a control group: Clear guidance is needed. *Contemp. Clin. Trials* **2022**, *113*, 106656. [[CrossRef](#)] [[PubMed](#)]
10. Rubin, M. When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese* **2021**, *199*, 10969–11000. [[CrossRef](#)]
11. García-Pérez, A. Use and misuse of correction for multiple testing. *Meth. Psych. B* **2023**, *8*, 100120. [[CrossRef](#)]
12. Li, G.; Taljaard, M.; Van den Heuvel, E.; Levine, M.; Cook, D.; Wells, G.; Devereaux, P.; Thabane, L. An introduction to multiplicity issues in clinical trials: The what, why, when and how. *Int. J. Epidemiol.* **2017**, *46*, 746–755. [[CrossRef](#)] [[PubMed](#)]
13. Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **1988**, *75*, 800–802. [[CrossRef](#)]
14. Kabe, D. On the exact distribution of the Fisher-Behrens’s-Welch statistic. *Metrika* **1996**, *10*, 13–15. [[CrossRef](#)]
15. Lepage, Y. A combination Wilcoxon’s and Ansari-Bradley’s statistics. *Biometrika* **1971**, *58*, 213–217. [[CrossRef](#)]
16. Bolthausen, E. Exact convergence rates in some Martingale central limit theorems. *Ann. Prob.* **1982**, *3*, 672–688.
17. Lee, S.; Cha, S.; Kim, I. General framework for conditional two-sample testing. *arXiv* **2024**. [[CrossRef](#)]
18. Yan, J.; Li, Z.; Zhang, X. Distance and kernel measures for global and local two-sample conditional distribution testing. *arXiv* **2024**. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.