International Journal of Health Sciences
December 2023, Vol. 11, No. 2, pp. 51-60
ISSN: 2372-5060 (Print), 2372-5079 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/ijhs.v11n2a7
URL: https://doi.org/10.15640/ijhs.v11n2a7

Assessing the Treatment Effect Based on Percentage Change from Baseline in a Controlled Trial - A Simulation Study

Sarfaraz Sayyed[1] (M.Sc.), Ashwini Mathur[2](Ph.D.) & Asha Kamath[3] (Ph.D.)

Abstract

Background: Several possibilities exist for analyzing continuous endpoints in randomized clinical trials. These include regressing post-treatment response, absolute change from baseline, or percent change from baseline on factors (gender, region, etc.) with/without baseline response as a covariate. If response variable follows a Gaussian distribution, the percent change from baseline will be the ratio of two correlated Gaussian distributions. The assumption that percent change from baseline follows a Gaussian distribution may be incorrect and biased. Additionally, missing data could complicate the behavior of the percent change variable. It is also shown by Vickers (Vickers, 2001) that percentage changes from baseline are statistically inefficient when analyzed traditionally.
Methods: We propose an alternative solution using the Delta method to get estimates under different missing data imputation techniques and investigate the distribution for percent change from baseline for all values in numerator and denominator except zero.
Results: Delta method estimates on simulated data were compared with traditional point estimates with confidence intervals.
Conclusions: The Proposed method provides results that are better, and this study would be useful to researchers in choosing methods for analysis and decision-making when the endpoint of interest is the ratio of correlated Gaussian distribution, and the data has missing responses.

Keywords: Ratio; Bivariate distribution; Gaussian distribution; Missing data; Multiple imputation; Delta method; Percent change from baseline.

1. Introduction

Using Percent change from baseline for continuous endpoints in clinical trials is quite common. Consider the percent change for continuous outcome from endpoint to baseline in clinical trials, the measurement before start of treatment administration (Y_1) and at the end of treatment period (Y_2) are assumed to be following Gaussian distribution with pdf (William, Matthew, 1984)

$$(Y_1, Y_2) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

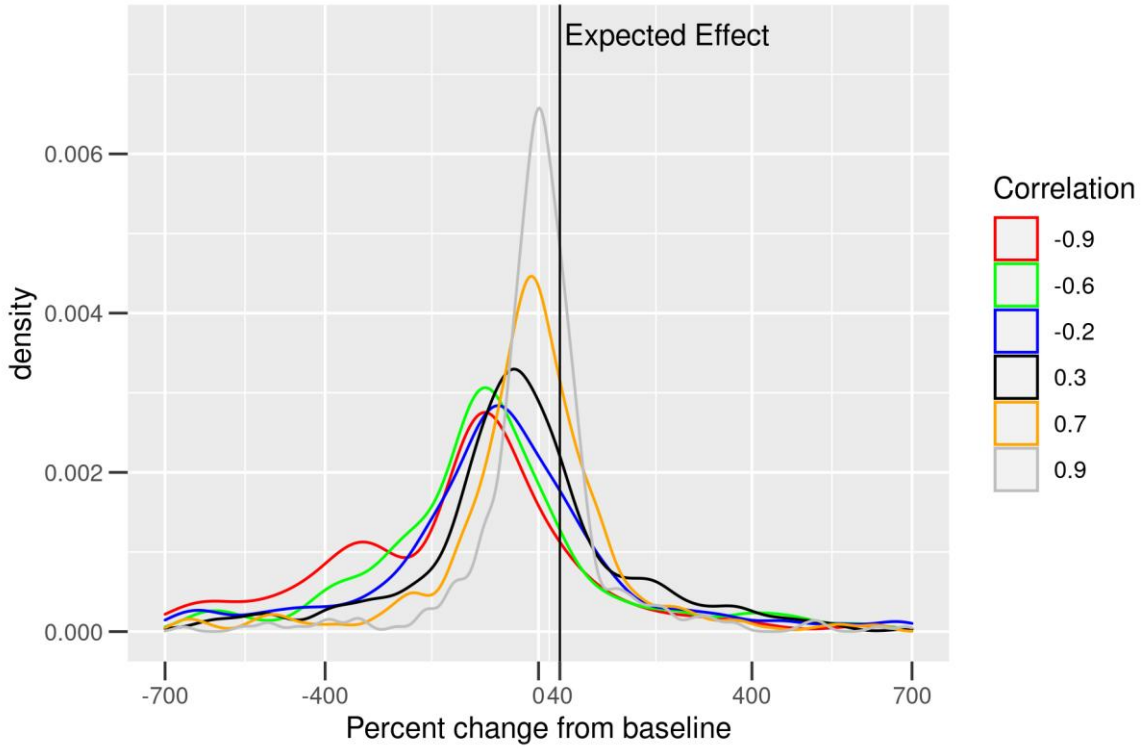
$$P(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \times \left\{\frac{(Y_1-\mu_1)^2}{\sigma_1^2} + \frac{(Y_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1,\sigma_2}\right\}\right) \quad (1)$$

where μ_i and σ_i are the mean and SD for Y_i i.e., for i^{th} time point and ρ is the correlation between Y_1 and Y_2

We simulated of sample size 800 (sample size not chosen with specific power) for Bivariate Gaussian distribution function (pdf) in equation (1) under each of the following correlations -0.9, -0.6, -0.2, 0.3, 0.7, 0.9 with unit standard deviation for both variables and the mean vector as [0.5, 0.7]. This means that the true percent change is (0.7-0.5) x 100/0.5 which is 40%. The density distribution and QQ-plots are given in Figure 1 and Figure 2 and the summary statistics for varying correlation coefficient embedded in Figure 2.

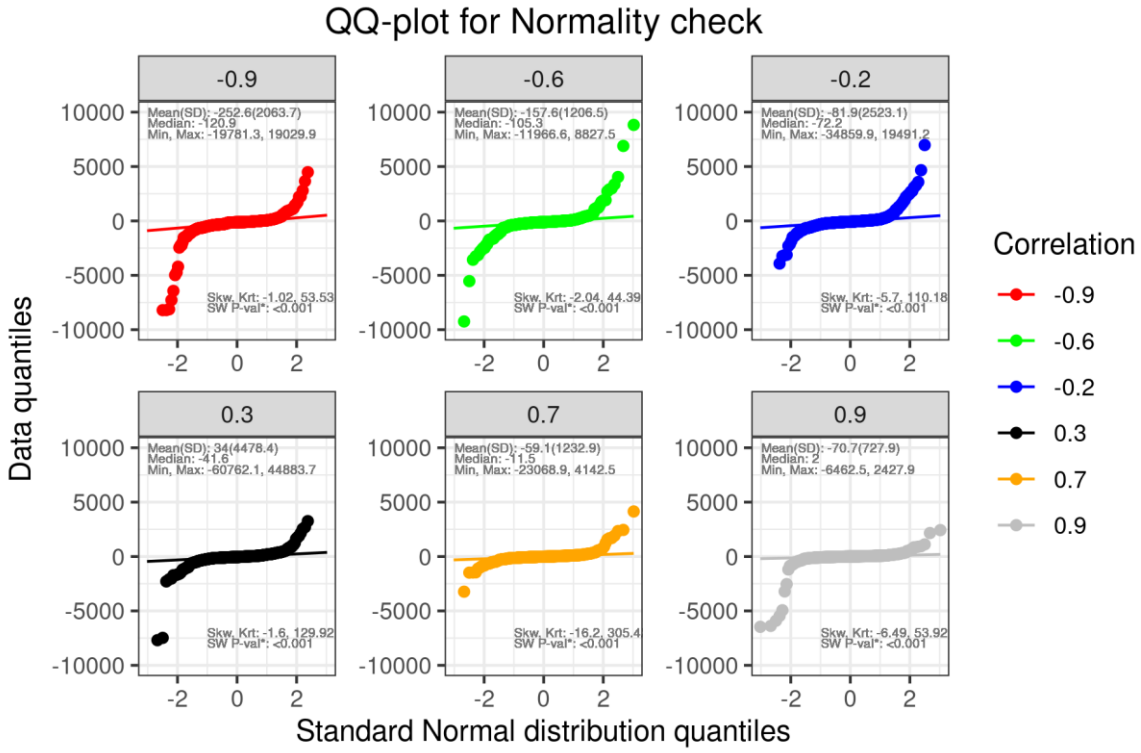
[1] Analytics, Novartis Healthcare Pvt. Ltd., Hyderabad, Telangana, India, E-mail: sarfaraz.sayyed@novartis.com
[2] Global Development Operations, Novartis Ireland Ltd., Dublin, Ireland, E-mail: ashwini.mathur@novartis.com
[3] Dept. of Data Science, Manipal academy of higher education, Manipal, Karnataka, India, E-mail: asha.kamath@manipal.edu

Figure 1: Distribution for percent change from baseline under differing correlations



The x-axis is restricted between (-700, 700) for visualization purpose.

Figure 2: QQ-plot for checking the normality assumption on the percent change from baseline variable



*P-value from Shapiro–Wilk normality test. SD: Standard deviation, Skw: Skewness, Krt: Kurtosis.
The y-axis for the qq-plot is restricted between (-10000, 10000) to have better visuals.

In distribution plot Figure 1, we can clearly observe the bias as the distribution is not centered around 40% which is the true percent change. When we see the QQ-plots, it clearly tells us that the data distribution is skewed with long tails which indicates that it is not normally distributed.

Skewness for a normal distribution or any symmetric data is zero whereas negative or positive values indicate a skewed distribution. We do not see a consistent pattern for different correlations, but skewness is non-zero for any value of correlation. Also, Kurtosis is far from 3 indicating heavy tails in the distribution of percent change from baseline. Moreover, the Shapiro Wilk's test for normality with statistically significant p-values also indicates that percent change data does not follow a Gaussian distribution for any of the correlations as appears in Figure 2.

The primary goal of this research was to find an appropriate way to estimate the treatment effect and its confidence interval when the endpoint of interest is percent change form baseline which is primarily based on ratio of correlated Gaussian distributions.

Methods

1.1 Notation

Let the outcome of interest be Y_{ijk} at jth (j = 1, 2) time point for the ith patient receiving kth (k =t, r) treatment, k=t refers to test drug and k=r refers to reference drug also j=1 refers to baseline and j=2 refers to post-baseline. The percent change from baseline in kth arm for the ith patient will then be given as

$$\frac{Y_{i2k} - Y_{i1k}}{Y_{i1k}} \text{x } 100 = (X_{ik} - 1) * 100, \qquad Where\ X_{ik} = \frac{Y_{i2k}}{Y_{i1k}}$$

Let $\mu_{1t}\&\mu_{2t}$ be the population mean and $\sigma_{1t}\&\sigma_{2t}$ be the standard deviation at baseline and end of treatment period respectively for test drug with correlation ρ_t.On similar lines let $\mu_{1r}\&\mu_{2r}$ be the population mean and $\sigma_{1r}\&\sigma_{2r}$ be the standard deviation at baseline and end of treatment period respectively for comparator drug with correlation ρ_r. Y_{ijk} follows bivariate Gaussian distribution.

1.2 Framework

Consider the ratio of post baseline vs baseline as below, the ratio 'X' would be area of

$$X = \frac{Y_2}{Y_1}$$

focus as adding or subtracting to calculate percent change from baseline using this ratio will just shift the location and multiplying will just change the scale, but the characteristic of the distribution will remain the same. The exact distribution given by Fieller (Fieller, 1932) for ratio $x$of correlated Gaussian distribution variables is

$$f(x) = \frac{b(x)d(x)}{\sqrt{(2\pi)\sigma_1\sigma_2 a^3(x)}}\left[\Phi\left\{\frac{b(x)}{\sqrt{(1-\rho^2)a(x)}}\right\} - \Phi\left\{\frac{-b(x)}{\sqrt{(1-\rho^2)a(x)}}\right\}\right] + \frac{\sqrt{(1-\rho^2)}}{\pi\sigma_1\sigma_2 a^2(x)}\ exp\left\{\frac{-c}{2(1-\rho^2)}\right\}, \quad (2)$$

where

$$a(x) = \left(\frac{x^2}{{\sigma_2}^2} - \frac{2\rho x}{\sigma_1\sigma_2} + \frac{1}{{\sigma_1}^2}\right)^{1/2}, \qquad b(x) = \left(\frac{\mu_2 x}{{\sigma_2}^2} - \frac{\rho(\mu_2 + \mu_1 x)}{\sigma_1\sigma_2} + \frac{\mu_1}{{\sigma_1}^2}\right),$$

$$c = \left(\frac{{\mu_2}^2}{{\sigma_2}^2} - \frac{2\rho(\mu_1 * \mu_2)}{\sigma_1\sigma_2} + \frac{{\mu_1}^2}{{\sigma_1}^2}\right), \quad d(x) = exp\left\{\frac{b^2(x) - c.a^2(x)}{2(1-\rho^2)a^2(x)}\right\}, \quad \Phi(h) = \int_{-\infty}^{h} \frac{1}{\sqrt{(2\pi)}}.e^{-\frac{1}{2}u^2}\,du$$

where,

$\mu_2\&\mu_1$: are the means at post baseline and baseline respectively.

$\sigma_2\&\sigma_1$: are the standard deviations at post baseline and baseline respectively.

ρ is the correlation between post baseline and baseline timepoints.

$(1-\alpha)$confidence interval (m_l, m_u)proposed by Fieller (Fieller, 1954) for the ratio x using above notations is given by

$$(m_l, m_u) = \frac{\left(Y_1 Y_2 - t_{r,\alpha}^2\rho\sigma_1\sigma_2\right) \mp \left\{\left(Y_1 Y_2 - t_{r,\alpha}^2\rho\sigma_1\sigma_2\right)^2 - \left(Y_1^2 - t_{r,\alpha}^2\sigma_1^2\right)\left(Y_2^2 - t_{r,\alpha}^2\sigma_2^2\right)\right\}}{Y_1^2 - t_{r,\alpha}^2\sigma_1^2}$$

This can also be expressed as,

$$(m_l, m_u) = \frac{1}{(1-g)}\left[\left(x - g\rho\frac{\sigma_2}{\sigma_1}\right) \pm \frac{t_{r,\alpha}}{Y_1}\sqrt{\sigma_2^2 - 2x\rho\sigma_1\sigma_2 + x^2\sigma_1^2 - g(\sigma_2^2 - \rho^2\sigma_2^2)}\right], \quad (2.1)$$

Where $g = \frac{\sigma_1 t_{r,\alpha}^2}{Y_1^2}$ & $t_{r,\alpha}$ is students t-distribution with r degrees of freedom & α level of significance.

1.3 Approximation to ratio Distribution

Hinkley (Hinkley, 1969) has given an approximation to the exact ratio distribution as below under the assumption that the denominator is always greater than zero and that coefficient of variation for the denominator tends towards 0.

$$f*(x) = \frac{b(x)\,d(x)}{\sqrt{2\pi\sigma_1\sigma_2 a^3(x)}} \tag{3}$$

where the definition of $a(x)$, $b(x) and d(x)$ remain same as defined equation (2).

Later Hinkley proposed correction (Hinkley, 1970) based on Marsaglia's (Marsaglia, 1965) pointing out that the general ratio can be expressed as

$$f*(x) = \frac{\sigma_2}{\sigma_1}\sqrt{1-\rho^2}\left\{\frac{\frac{\mu_2}{\sigma_2} - Y_2}{\frac{\mu_1}{\sigma_1} - Y_1} + \frac{\rho}{\sqrt{1-\rho^2}}\right\} \tag{4}$$

Marsaglia (Marsaglia, 1965) has shown that the correlated variables Y_2 / Y_1 can be expressed in the standard form as $\frac{a+r}{b+s}$, where a, b ≥ 0 are constants with r and s being independent standard gaussian variates. He pointed out that the ratio $\frac{a+r}{b+s}$ follows approximate gaussian distribution for (a, b) = (< 2.256, > 4) with

$$mean\ \mu = \frac{a}{(1.01 * b) - 0.2713}\ and\ variance\ \sigma^2 = \frac{a^2+1}{(b^2 + (0.108 * b) - 3.795)} - \mu^2$$

Hayyas, et al., in their work (Jack, Donald, Nicolas, 1975) came up with normal approximation for the ratio of two gaussian variables which are not necessarily statistically independent under the conditions that the correlation is ≤ 0.5, the coefficient of variation for denominator variable < 0.09 and that for numerator variable is > 0.19. They also claimed that Geary-Hinkley normal approximation to the transformation is robust if coefficient of variation for the denominator variable < 0.39 and that of numerator variable is >0.005.

Geary (Geary, 1930) proposed a transformation to achieve normal approximation for the ratio under the condition that the denominator has a small coefficient of variation.

Most recent work was proposed by Saralees (Saralees, 2006) who derived the distributions of the ratio's when the joint distribution of the numerator and denominator variables are either "elliptically symmetric Pearson-type II distribution" or "elliptically symmetric Pearson-type VII distribution" or "elliptically symmetric Kotz-type distribution". Under the "elliptically symmetric Kotz-type distribution" with specific values to some parameters this type would reduce to a bivariate normal distribution.

1.4 First and Second order Moments

Getting closed form of the moments for the distributions suggested in equation (3), equation (4) and other authors is difficult to obtain. Also, the integral for these distributions doesn't converge for all the combination of ratios which is also highlighted in the above section.

1.5 Delta Method

If distribution of random variable is not specifically of interest but rather the interest is in an estimate of a function of this random variable, Delta method (Casella, Berger, 2002) is helpful. In situations, as we have for percent change from baseline or ratio of random variables, the distribution is complex. But our interest is in estimating the functions of these variables rather than the distribution. The delta method theorem states that if we have a sequence of random variables,

$Y = (Y_1, Y_2, \ldots, Y_n)\ with$

$$E(Y_1, Y_2, \ldots, Y_n) = E(Y) = \mu = (\mu_1, \mu_2, \ldots, \mu_n)\ and\ Cov(Y_i, Y_j) = \sigma_{ij}$$

Then for given function "g" for specific value of μ and existing nonzero continuous first order partial derivative $g'(\mu)$ for which

$$\xi^2 = \sum_i \sum_j \sigma_{ij}\, g_i'(\mu) g_j'(\mu) > 0, then$$

$$\sqrt{n}(g(\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_s)) - g(\mu_1, \mu_2, \ldots, \mu_s)) \rightarrow N(0, \xi^2)\ in\ distribution$$

1.6 Estimate of ratio using Delta Method

When dealing with ratio of two correlated variables $(Y_1,\ Y_2)$ the parametric function to estimate will be $g(\mu_1,\mu_2) = \mu_2/\mu_1$. The first order derivatives for parametric function $g(\mu_1,\mu_2)$ will be

$$\frac{\partial g(\mu_1,\mu_2)}{\partial \mu_2} = \frac{1}{\mu_1}\ ,\quad \frac{\partial g(\mu_1,\mu_2)}{\partial \mu_1} = \frac{-\mu_2}{\mu_1^2}\ and$$

$$\xi^2 = \sum_i \sum_j \sigma_{ij}\, g_i'(\mu) g_j'(\mu) = \frac{1}{\mu_1^2}\sigma_2^2 + \frac{\mu_2^2}{\mu_1^4}\sigma_1^2 - 2\frac{\mu_2}{\mu_1^3}\sigma_{12}$$

So, the estimate for the mean and variance of the correlated ratio will be given as

$$E\left(\frac{Y_2}{Y_1}\right) = \frac{\bar{Y}_2}{\bar{Y}_1} \approx \frac{\mu_2}{\mu_1}\ and \tag{5}$$

$$n\,.Var\left(\frac{Y_2}{Y_1}\right) \approx \frac{1}{\mu_1^2}\sigma_2^2 + \frac{\mu_2^2}{\mu_1^4}\sigma_1^2 - 2\frac{\mu_2}{\mu_1^3}\sigma_{12} \tag{6}$$

For variance, the components will be estimated from the data as

$\widehat{\mu_1} = \bar{Y}_1,\quad \widehat{\mu_2} = \bar{Y}_2,\ \widehat{\sigma_1} = sample\ SD\ for\ Y_1 = s_1,\ \widehat{\sigma_2} = sample\ SD\ for\ Y_2 = s_2,$

$\widehat{\sigma_{12}} = sample\ covariance\ between\ Y_1 and\ Y_2 = s_{12}$

1.7 Estimate of treatment difference for % change from baseline

We will use the notations used in the Notation and Framework section.

Consider the percent change from baseline in the 2 groups

$$\frac{Y_{2k} - Y_{1k}}{Y_{1k}} \text{x } 100 = \left(\frac{Y_{2k}}{Y_{1k}} - 1\right) \text{x } 100 = (X_k - 1) \text{x } 100,$$

k takes value "t" or "r" for test group and reference group respectively.

Difference between the percent change from baseline in the two treatment groups will be given by

$$\big((X_t - 1) \text{ x } 100\big) - \big((X_r - 1) \text{ x } 100\big) = (X_t - X_r) \text{ x } 100 = \left[\frac{Y_{2t}}{Y_{1t}} - \frac{Y_{2r}}{Y_{1r}}\right] \text{x} 100$$

So, the parametric function we need to estimate will be

$$G = \left[\frac{\mu_{2t}}{\mu_{1t}} - \frac{\mu_{2r}}{\mu_{1r}}\right] \text{x} 100$$

Using Delta method as stated in equation (5) we get,

$$E\left(\frac{Y_{2t}}{Y_{1t}} - \frac{Y_{2r}}{Y_{1r}}\right) = E\left(\frac{Y_{2t}}{Y_{1t}}\right) - E\left(\frac{Y_{2r}}{Y_{1r}}\right) = \frac{\bar{Y}_{2t}}{\bar{Y}_{1t}} - \frac{\bar{Y}_{2r}}{\bar{Y}_{1r}} \approx \frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{22}}{\mu_{21}} \tag{7}$$

Since we are looking at the difference of 2 independent ratios the pooled variance for the estimate will be given by

$$Variance\ of\ \left(\frac{Y_{2t}}{Y_{1t}} - \frac{Y_{2r}}{Y_{1r}}\right) = V\left(\frac{Y_{2t}}{Y_{1t}} - \frac{Y_{2r}}{Y_{1r}}\right) = V\left(\frac{Y_{2t}}{Y_{1t}}\right) + V\left(\frac{Y_{2r}}{Y_{1r}}\right), \tag{8}$$

Now equation (8) is sum of variance of ratio's which can be calculated using equation (6).

1.8 Estimate of treatment difference for % change from baseline in presence of covariate

We will use the notations used in the Notation and Framework section.Consider the ratio 'X' for post-baseline vs baseline in each group to be a response variable. Now to get the least squares estimate in the presence of covariates, we get the least squares estimator function and find the delta method estimate for it. For e.g., consider the following model:

$$X_{kci} = \mu + \alpha_1 D + \alpha_2 C + \alpha_3 DC + \varepsilon'_{kci} \tag{9}$$

Where X_{kci} is the response from ith subject receiving kth treatment at cth center,

D = treatment (Test or reference), we are here considering only two treatments in the design.

C = center effect (center 1, center 2, ...),

DC = interaction between Drug and center and

ε'_{kci} is the error term.

Here X_{kc} is the ratio of Y_{2kc}&Y_{1kc} i.e., post vs pre-baseline assessment.

So, the estimated mean for the model in equation (9) will be given by

$$E(X_{kc}) = \mu + \alpha_1 D + \alpha_2 C + \alpha_3 DC \tag{10}$$

The least squares mean (Lsmean) for the comparator group will be obtained by averaging over the levels of the covariate using the equation (10).

So, the comparator drug Lsmean in the presence of covariate will be given by,

$$E(X_r) = \left(E(X_{r1}) + E(X_{r2}) + \cdots + E(X_{rp})\right)/p \quad (11)$$

Where $E(X_{rc})$is the expected effect in terms of ratio of post baseline vs pre baseline for reference drug at center 'c' having 'p' centers.

Similarly, the Test drug Lsmean in the presence of covariate will be given by,

$$E(X_t) = \left(E(X_{t1}) + E(X_{t2}) + \cdots + E(X_{tp})\right)/p \quad (12)$$

Where $E(X_{tc})$is the expected effect in terms of ratio of post baseline vs pre baseline for test drug at center 'c' having 'p' centers.

And hence the investigational test drug effect is the difference in the Lsmeans obtained in equation (11) and equation (12) given by,

$$E(X_t) - E(X_r) = \frac{1}{p}\sum_{c=1}^{p}\left(E(X_{tc}) - E(X_{rc})\right), \qquad where\ p\ is\ total\ number\ of\ centers \quad (13)$$

Using Delta method estimate as derived in equation (5) for equation (13) we get,

$$E(X_t) - E(X_r) = \frac{1}{p}\sum_{c=1}^{p}\left(E(X_{tc}) - E(X_{rc})\right) = \frac{1}{p}\sum_{c=1}^{p}\left(E\left(\frac{Y_{2tc}}{Y_{1tc}}\right) - E\left(\frac{Y_{2rc}}{Y_{1rc}}\right)\right)$$

$$= \frac{1}{p}\sum_{c=1}^{p}\left(\left(\frac{\bar{Y}_{2tc}}{\bar{Y}_{1tc}}\right) - \left(\frac{\bar{Y}_{2rc}}{\bar{Y}_{1rc}}\right)\right) \quad (14)$$

Where, $\bar{Y}_{2kc}$ & $\bar{Y}_{1kc}$ *are mean for* k^{th} *drug at center c for post baseline & baseline respectively.*

Variance for this estimate will be given by

$$V\left(\frac{1}{p}\sum_{c=1}^{p}\left(\left(\frac{Y_{2tc}}{Y_{1tc}}\right) - \left(\frac{Y_{2rc}}{Y_{1rc}}\right)\right)\right) = \frac{1}{p^2}\left(\sum_{c=1}^{p}\left(V\left(\frac{Y_{2tc}}{Y_{1tc}}\right) + V\left(\frac{Y_{2rc}}{Y_{1rc}}\right)\right)\right) \quad (15)$$

Now since equation (15) is sum of variance of ratio's which can easily be calculated using equation (8). We obtain the Lsmean and its standard error (SE) using both the methods. Under Delta method Results are averaged over the levels of the covariate in both the arms and then the difference between the two arms is obtained.

1.9 Expressing difference of 2 ratios as one ratio

Consider the post-baseline and baseline variables Y_{2k} *and* Y_{1k} respectively with k=t for test drug and k=r for reference drug. Consider the difference in percent change from baseline for the 2 groups;

$$\left(\frac{Y_{2t} - Y_{1t}}{Y_{1t}}\text{x } 100\right) - \left(\frac{Y_{2r} - Y_{1r}}{Y_{1r}}\text{x } 100\right) = \left(\left(\frac{Y_{2t}}{Y_{1t}} - 1\right)\text{x } 100\right) - \left(\left(\frac{Y_{2r}}{Y_{1r}} - 1\right)\text{x } 100\right)$$

$$= \left(\frac{Y_{2t}}{Y_{1t}} - \frac{Y_{2r}}{Y_{1r}}\right) \times 100 = \left(\frac{Y_{2t}\,Y_{1r} - Y_{2r}\,Y_{1t}}{Y_{1t}\,Y_{1r}}\right) \times 100$$

Keeping the constant multiplier aside which can be adjusted later with the estimates, the ratio $\left(\frac{Y_{2t}\,Y_{1r} - Y_{2r}\,Y_{1t}}{Y_{1t}\,Y_{1r}}\right)$ can be considered as single ratio of 2 correlated variates and using Fieller's (1-α) confidence limits in equation (2.1) we can drive the confidence limits for this ratio. We need to find the variance covariance matrix for numerator and denominator, the expression for which is derived below.

Variance of Numerator

$$V(Y_{2t}\,Y_{1r} - Y_{2r}\,Y_{1t}) = V(Y_{2t}\,Y_{1r}) + V(Y_{2r}\,Y_{1t}) - 2 \times Cov(Y_{2t}\,Y_{1r}, Y_{2r}\,Y_{1t})$$

where,

$$V(Y_{2t}\,Y_{1r}) = V(Y_{1r})E(Y_{2t})^2 + V(Y_{2t})E(Y_{1r})^2 + V(Y_{1r})V(Y_{2t}),$$

$$V(Y_{2r}\,Y_{1t}) = V(Y_{2r})E(Y_{1t})^2 + V(Y_{1t})E(Y_{2r})^2 + V(Y_{2r})V(Y_{1t}),$$

$$Cov(Y_{2t}\,Y_{1r}, Y_{2r}\,Y_{1t}) = Cov(Y_{2t}, Y_{1t})Cov(Y_{2r}, Y_{1r}) + Cov(Y_{2t}, Y_{1t})E(Y_{2r})E(Y_{1r}) + Cov(Y_{2r}, Y_{1r})E(Y_{2t})E(Y_{1t})$$

Variance of Denominator

$$V(Y_{1t}\,Y_{1r}) = V(Y_{1r})E(Y_{1t})^2 + V(Y_{1t})E(Y_{1r})^2 + V(Y_{1r})V(Y_{1t}),$$

Covariance between Numerator and Denominator

$$Cov(Y_{2t}Y_{1r} - Y_{2r}Y_{1t}, Y_{1t}Y_{1r}) = Cov(Y_{2t}Y_{1r}, Y_{1t}Y_{1r}) - Cov(Y_{2r}Y_{1t}, Y_{1t}Y_{1r})$$

where,

$$Cov(Y_{2t}Y_{1r}, Y_{1t}Y_{1r}) = Cov(Y_{2t}, Y_{1t})[V(Y_{1r}) + E(Y_{1r})^2] + V(Y_{1r})E(Y_{2t})E(Y_{1t}),$$

$$Cov(Y_{2r}Y_{1t}, Y_{1t}Y_{1r}) = Cov(Y_{2r}, Y_{1r})[V(Y_{1t}) + E(Y_{1t})^2] + V(Y_{1t})E(Y_{2r})E(Y_{1r})$$

1.10 Simulation with non-response and non-compliance

We started with a sample size of 800 subjects (400 subjects per arm) for obtaining treatment difference of around 0.72 units to be simulated from Multivariate Gaussian distribution. We also introduced artificial missingness in the data to observe the behavior of the methods on the estimate. The missingness introduced in simulated data was roughly 10%, 15%, 20% and 25%. Subjects having post baseline assessment below median were randomly chosen to go missing with desired percent to reflect a scenario for missing due to lack of efficacy.

The simulation was repeated 1000 times to reflect the robustness of the methods. For calculating Delta method estimate and SE in Table 1, equation (7) and equation (8) were used.

For addressing missing data, we used 3 widely used techniques in clinical trials (Blankers, Koeter, Schippers, 2010),

1) "Complete case analysis"(CC): Here all the rows with complete data were used and the rows with missing data were ignored.

2) "Last observation carried forward" (LOCF): Here the missing data was imputed using baseline data for the same id and then the complete data was used for analysis.

3) "Multiple imputation" (MI): Here multiple imputation technique was used for imputing the missing data which created multiple complete datasets. Each time MI was run to create 10 copies of compete datasets each with 50 iterations. Each of these datasets were analyzed separately and the results from each dataset were pooled using Rubin's rule (Rubin, 1987).

2. Results

Table 1: Simulation results taking into consideration correlation=0.8 without Covariate

	Estimates for percent change from baseline using linear model			Estimates using Delta method		
Data missing %	Estimate (SE)	95% CI	Bias	Estimate (SE)	95% CI	Bias
	0.53(0.25)	(0.04,1.01)	0.189	0.72(0.24)	(0.24,1.19)	0.001
Complete Case						
10	0.58(0.26)	(0.06,1.09)	0.142	0.76(0.24)	(0.29,1.24)	0.046
15	0.60(0.27)	(0.08,1.13)	0.113	0.79(0.24)	(0.32,1.26)	0.073
20	0.63(0.27)	(0.1,1.17)	0.084	0.82(0.24)	(0.35,1.28)	0.099
25	0.67(0.28)	(0.13,1.21)	0.047	0.85(0.23)	(0.39,1.30)	0.131
Last observation carried forward						
10	0.52(0.24)	(0.06,0.98)	0.199	0.69(0.23)	(0.24,1.14)	0.03
15	0.51(0.23)	(0.07,0.96)	0.203	0.67(0.22)	(0.24,1.11)	0.044
20	0.51(0.22)	(0.08,0.93)	0.211	0.66(0.21)	(0.24,1.07)	0.061
25	0.50(0.21)	(0.1,0.91)	0.215	0.64(0.20)	(0.25,1.04)	0.075
Multiple Imputation						
10	0.56(0.27)	(0.02,1.1)	0.155	0.75(0.25)	(0.25,1.24)	0.03
15	0.58(0.28)	(0.03,1.14)	0.133	0.77(0.26)	(0.26,1.27)	0.049
20	0.61(0.29)	(0.04,1.19)	0.105	0.79(0.26)	(0.29,1.30)	0.073
25	0.64(0.30)	(0.06,1.22)	0.073	0.82(0.26)	(0.32,1.32)	0.1

The above exercise was also repeated by adding a covariate in the model. We chose here a binary type of categorical covariate and kept the treatment effect around 0.79%. We fitted the model with percent change from baseline as dependent variable with treatment and covariate as independent variables and obtained the Lsmean with its SE. The estimate and its SE for Delta method in Table 2 were calculated using equation (14) and equation (15). For both the methods absolute bias and 95% confidence interval (95% CI, mean ± 1.96*SE) is provided. This was done with the same sample size of 400 per treatment group.

Table 2: Simulation results taking into consideration correlation=0.8 with Covariate

	Estimates for percent change from baseline using linear model			Estimates using Delta method		
Data missing %	Lsmean Estimate (SE)	95% CI	Bias	Lsmean Estimate (SE)	95% CI	Bias
	0.54(0.52)	(-0.47,1.56)	0.249	0.79(0.49)	(-0.17,1.75)	0.001
Complete Case						
10	0.55(0.55)	(-0.53,1.63)	0.237	0.79(0.49)	(-0.16,1.75)	0.005
15	0.55(0.57)	(-0.58,1.67)	0.243	0.78(0.49)	(-0.18,1.74)	0.008
20	0.56(0.59)	(-0.60,1.72)	0.229	0.79(0.49)	(-0.17,1.75)	0.003
25	0.57(0.62)	(-0.65,1.78)	0.224	0.80(0.49)	(-0.16,1.75)	0.007
Last observation carried forward						
10	0.50(0.96)	(-1.38,2.37)	0.292	0.75(0.93)	(-1.07,2.56)	0.042
15	0.46(1.06)	(-1.61,2.53)	0.325	0.73(1.02)	(-1.28,2.73)	0.064
20	0.45(1.11)	(-1.72,2.62)	0.34	0.71(1.08)	(-1.40,2.82)	0.083
25	0.43(1.12)	(-1.78,2.63)	0.364	0.68(1.09)	(-1.45,2.82)	0.105
Multiple Imputation						
10	0.51(0.54)	(-0.55,1.57)	0.279	0.74(0.48)	(-0.21,1.68)	0.051
15	0.47(0.55)	(-0.61,1.54)	0.322	0.71(0.48)	(-0.23,1.64)	0.084
20	0.44(0.56)	(-0.66,1.54)	0.348	0.67(0.48)	(-0.26,1.61)	0.117
25	0.40(0.57)	(-0.72,1.51)	0.391	0.63(0.47)	(-0.30,1.56)	0.158

In Table 2, due to the addition of the covariate, the sample size of 400 per treatment group is not enough to detect the true difference of 0.72 as all confidence intervals include 0. We ran additional simulation with sample size of 800 per treatment group and results are given in Table 3 below.

Table 3: Simulation results taking into consideration correlation=0.8 with Covariate and 800 subjects

	Estimates for percent change from baseline using linear model			Estimates using Delta method		
Data missing %	Lsmean Estimate (SE)	95% CI	Bias	Lsmean Estimate (SE)	95% CI	Bias
	0.54 (0.37)	(-0.18,1.26)	0.252	0.79 (0.35)	(0.11,1.47)	0.001
Complete Case						
10	0.54 (0.39)	(-0.22,1.31)	0.248	0.79 (0.35)	(0.11,1.47)	0.002
15	0.55 (0.41)	(-0.25,1.34)	0.243	0.79 (0.35)	(0.11,1.47)	0.001
20	0.55 (0.42)	(-0.27,1.37)	0.239	0.79 (0.35)	(0.11,1.47)	0.001
25	0.55 (0.44)	(-0.31,1.41)	0.237	0.79 (0.35)	(0.11,1.47)	0.002

Last observation carried forward						
10	0.49 (0.68)	(-0.84,1.82)	0.302	0.75 (0.66)	(-0.54,2.03)	0.044
15	0.46 (0.75)	(-1,1.93)	0.325	0.72 (0.72)	(-0.70,2.14)	0.066
20	0.44 (0.78)	(-1.10,1.98)	0.349	0.70 (0.76)	(-0.79,2.19)	0.087
25	0.41 (0.79)	(-1.14,1.97)	0.375	0.68 (0.77)	(-0.83,2.19)	0.109
Multiple Imputation						
10	0.50 (0.38)	(-0.25,1.25)	0.293	0.71 (0.34)	(0.04,1.37)	0.081
15	0.47 (0.39)	(-0.29,1.23)	0.32	0.71 (0.34)	(0.04,1.37)	0.081
20	0.43 (0.40)	(-0.35,1.21)	0.357	0.71 (0.34)	(0.04,1.37)	0.0806
25	0.39 (0.41)	(-0.40,1.19)	0.397	0.71 (0.34)	(0.04,1.37)	0.0806

In all the tables (Table 1, Table 2, Table 3) with simulation results we have added a column for absolute bias ignoring the direction of bias for both delta method as well as linear model estimates. The bias represents the difference between the observed and expected treatment effects and was calculated by subtracting the expected difference between the groups from the observed difference between the groups.

3. Discussion

In Table 1, Table 2 and Table 3 delta method clearly shows significantly lesser bias and smaller standard error when the end point is percent change from baseline as compared to calculating percent change for each subject and fitting linear model. In case of missingness in the data, delta method picks up significance for CC and MI methods compared to traditional method. LOCF method is least powerful in detecting the significance. Proposed method from the above tables shows delta method to perform better then calculating percent change for each patient as response and then performing linear modelling. For missing data, delta method using complete case analyses or Multiple imputation should be preferred over LOCF method. We have only considered the case with linear relationship; further exploration might be required to see the performance of the method in other settings.

We also tried to calculate the confidence intervals for the difference between two drugs on percent change from baseline as endpoint using Fieller's $(1 - a)$ confidence for ratio (Fieller,1954). The idea was to express the difference in 2 ratios as a single ratio, get the variance covariance matrix (expression in Appendix) for this transformed ratio and then use the Fieller's $(1 - a)$ confidence interval for ratio specified in equation (2.1). The intervals obtained for running the simulation for complete data were (-10.269, 10.103) which were too wide because of the week correlation between the numerator & denominator and too high variance of the numerator. Delta method and Fieller's method both provide good estimates when the correlation between the numerator and denominator is high (Beyene, 2005), but Fieller's $(1 - a)$ confidence interval expression produces much wider length confidence interval when the correlation between numerator and denominator is week.

The sample size calculation for percent change from baseline having ratio of correlated variates is also a challenge which can be future scope of work in this area.

4. Availability of data and materials

Software available from: The Comprehensive R Archive Network (r-project.org) (version 4.2.1.)
Source code available from: https://github.com/Sarfaraz-Sayyed/Percent-change-from-baseline
Archived source code at time of publication: https://doi.org/10.5281/zenodo.7477421
License: creative commons License.

5. Acknowledgements

We would like to thank Novartis Healthcare Pvt. Ltd. and Manipal Academy of Higher Education for the support to carry out this research and all the reviewers for providing their valuable feedback and suggestions.

6. Funding

The author(s) declared that no grants were involved in supporting this work.

References

Berger., George Casella, & Roger L. (2002). *Statistical Inference* (Second Edition ed.). Duxbury Advanced series.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika, 24*, 428 - 440. doi:https://doi.org/10.2307/2331976

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B (Methodological), 16 (2)*, 175–185.

Geary, R. C. (1930). The Frequency Distribution of the Quotient of Two Normal Variates. *Journal of the Royal Statistical Society, 93*, 442-446. doi:https://doi.org/10.2307/2342070

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika, 56*, 635 – 639. doi:https://doi.org/10.2307/2334671

Hinkley, D. V. (1970, Decemeber). Correction: On the ratio of two correlated normal random variables. *Biometrika, 57*(3), 683. doi:https://doi.org/10.2307/2334796

Jack Hayyas, Donald Armstrong, & Nicolas Gressis. (1975, July). A Note on the ratio of two normally distributed variables. *Management Science, 21*(11). doi:https://doi.org/10.1287/mnsc.21.11.1338

Joseph Beyene, & Rahim Moineddin. (2005, October 12). Methods for confidence interval estimation of a ratio parameter with application to location quotients. *BMC Medical Research Methodology*, 5, Article number:32. doi:https://doi.org/10.1186/1471-2288-5-32

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Willey and Sons.

Marsaglia G. (1965). Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association, 60*, 193–204. doi:https://doi.org/10.2307/2283145

Matthijs Blankers, Maarten W J Koeter, & Gerard M Schippers. (2010, December 19). Missing Data Approaches in eHealth Research: Simulation Study and a Tutorial for Nonmathematically Inclined Researchers. *Journal of Medical Internet Research, 12, 5 e54*. doi:https://doi.org/10.2196/jmir.1448

Saralees Nadarajah. (2006). On the ratio X/Y for some elliptically symmetric distributions. *Journal of Multivariate Analysis, 97*, 342 - 358.

Vickers, A.J. (2001, June 28). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology*, 1, Article number:6. doi:https://doi.org/10.1186/1471-2288-1-6

William R. Dillon, & Matthew Goldstein. (1984). *Multivariate Analysis Methods and Applications.* New York; Chichester; Brisbane; Toronto; Singapore: John Willey & Sons, Inc.