



How to Analyze Change from Baseline: Absolute or Percentage Change?

Examiner:
Lars Rönnegård

Supervisor:
Johan Bring

Co-supervisor:
Richard Stridbeck

Author:
Ling Zhang
Kun Han

Date:
June 10, 2009

Högskolan Dalarna
781 88 Borlänge
Tel vx 023-778000

How to Analyze Change from Baseline: Absolute or Percentage Change?

June 10, 2009

ABSTRACT

In medical studies, it is common to have measurements before and after some medical interventions. How to measure the change from baseline is a common question met by researchers. Two of the methods often used are absolute change and percentage change. In this essay, from statistical point of view, we will discuss the comparison of the statistical power between absolute change and percentage change. What's more, a rule of thumb for calculation of the standard deviation of absolute change is checked in both theoretical and practical way. Simulation is also used to prove both the irrationality of the conclusion that percentage change is statistical inefficient and the nonexistence of the rule of thumb for percentage change. Some recommendations about how to measure change are put forward associated with the research work we have done.

KEY WORDS: *Absolute Change, Percentage Change, Baseline, Follow-up, Statistical Power, Rule of Thumb.*

1. Introduction

IN medical studies, a common way to measure treatment effect is to compare the outcome of interest before treatment with that after treatment. The measurements before and after treatment are known as the baseline (B) and the follow-up (F), respectively. How to measure the change from baseline is a common question met by researchers. There are many methods that can be used as the measure of difference. Two of them, which are used in a lot of clinical studies, are absolute change ($C = B - F$) and percentage change ($P = (B - F)/B$). In different books and articles, absolute change may also be called change, while percentage change is also called relative change.

There is a simple example that will show us the difference between absolute change and percentage change more clearly: Two obese men **A** and **B** participate in a weight loss program. Their weights at the beginning of the program are 150 Kg and 100 Kg, respectively. When they finish the program, the man **A** who weighs 150 Kg lost 15 Kg, while another man lost 10 Kg. From the example, we see that, the man **A** lost 5 Kg more than that the man **B**, but the percent of weight they lost are 10% in both cases. We want to know, which change measurement is best to show the treatment effect of the weight loss program.

In different clinical studies, either absolute change or percentage change may be chosen. In the study of healthy dieting and weight control, Waleekhachonloet (2007) used absolute

change to evaluate the change of weight. Neovius (2007) also chose absolute change as the change measurement in their obesity research, while Kim (2009) chose percentage change to measure the fat lost in different part of an obese man's body in a weight loss program. In a cystic fibrosis clinical study, Lavange (2007) used percentage change as well. We see that, both of the two methods have been used in different kinds of clinical studies.

The properties of absolute change and percentage change have been discussed by Tönqvist (1985). From his point of view, one of the advantages of percentage change is that percentage change is independent of the unit of measurement. For instance, a man who weighs 100 Kg lost 10% of weight after a treatment, i.e. 10 Kg. Equivalently, he lost 22.05 pounds (1Kg = 2.2046 Pounds). 10 Kg and 22.05 pounds are essentially the same weight, but the absolute change scores are different. However, no matter what the unit of measurement is, the percentage change is 10% all the time. More details about the advantage of absolute change can be found in the article of Tönqvist (1985). Although there are many advantages for the two change measurement methods, Tönqvist (1985) did not give any recommendations about how to make a choice between absolute change and percentage change based on these properties.

In other literatures, several suggestions about which method to choose are mentioned. Vickers (2001) suggested avoiding using percentage change. That is because he compared the statistical power of different methods by doing a simulation and concluded that percentage change from baseline is statistically inefficient. Kaiser (1989) also gave

some recommendations for making a choice between absolute change and percentage change. He suggested using the change measurement that has less correlation with baseline scores. A test statistic developed by Kaiser (1989) was also derived, i.e. the ratio of the maximum likelihood of absolute change to that of percentage change. The absolute change is recommended if the value of the test statistic is larger than one, while percentage change is preferred when it is less than one. That is, a simple rule helping researchers to make a choice quickly.

Actually, the primary consideration for choosing a change measurement method is different from different points of view. From a clinical point of view, we prefer to use a change measurement that may show the health-improvement for the patients in a more observable way. For example, in study of asthma, the primary outcome variable is often FEV (Forced Expiratory Volume L/s). The efficiency of a treatment is evaluated by calculating the percentage change in FEV from baseline. In hypertension studies, it is common to use the absolute change in blood pressure instead of percentage change. From statistical point of view, we prefer the method which has the highest statistical power as Vickers (2001) did.

Another issue concerned by researchers is the standard deviation of the treatment effect (change scores). For two medical interventions that may lead to the same expected change, the effect of the intervention that has a smaller standard deviation seems more stable and effective. And clinicians may always prefer that medical intervention. Since it is not practical for researchers to get all the interested experimental datasets which recorded the details of baseline and follow-up scores for each patient, there is a rule of thumb $SD(C) \approx SD(B) / \sqrt{2}$ ¹. It may help calculating the standard deviation of change scores from the standard deviation of baseline scores.

The aim of this essay is to show, the statistical efficiency of percentage change under some conditions in contrast with Vickers' (2001) conclusion, and the rationality of the rule of thumb. The essay is organized as follows. The second section is the comparison of the statistical power of absolute change and percentage change by constructing a test statistic under certain distribution assumption. In the third section, the rationality of rule of thumb is discussed in a theoretical way. Simulations of some of the issues discussed in section two and three are carried out in the fourth section. The fifth section is an empirical investigation of the usefulness of the rule of thumb by using some real datasets. Finally, in the discussion section, we discuss the results got from the previous sections, and give some suggestions.

2. Comparison of the statistical power of absolute change and percentage change

IN clinical research, it is common to test whether there is a treatment effect after a medical intervention. In order to test the treatment effect, it's necessary to choose a suitable measurement of the difference between baseline and follow-up scores.

From a statistical point of view, an important criterion for a good statistical method is high statistical power. Therefore, from the two common change measurement methods, absolute change and percentage change, the one with a higher statistical power will be preferred.

2.1 Statistical Power

According to the hypothesis testing theory, statistical power is the probability that a test reject the false null hypothesis.

The definition of statistical power can be expressed as equation (1).

$$\text{Statistical power} = P(\text{reject } H_0 \mid H_0 \text{ is False}) \quad (1)$$

where H_0 is the null hypothesis.

In a t-test, equation (1) can be rewritten as equation (2).

$$\begin{aligned} \text{Statistical power} &= P(\text{reject } H_0 \mid H_0 \text{ is False}) \quad (2) \\ &= P(|t| > t_{\alpha/2}) \\ &= P(p_t < \alpha) \end{aligned}$$

where $t_{\alpha/2}$ is the t-value under the significant level α in a two-side t-test, and p_t is the p-value of the t-test.

From expression (2), we may see that the larger the expected absolute value of the t-statistic is, the higher the statistical power will be.² Equivalently, the smaller the expected p-value is, the higher the statistical power will be. Therefore, from different measurement methods, we will choose the one that has a larger expected absolute value of t-statistic or a smaller expected p-value.

2.2 A Clinical Example of Blood Pressure Drug Experiment

To easily interpret the difference of two measurement methods, an example from a clinical trial is shown in Table 1. In the table, there are the records of the supine systolic blood pressures (in mmHg) for 5 patients before and after taking the drug captopril.

Let (B_j, F_j) denote a baseline/follow-up pair of scores for patient j in the treatment group, $j = 1, 2, \dots, n$. Then, we

¹Personal communication Prof. Johan Bring, E-mail: johan.bring@statisticon.se

²However, it must be emphasized that, in this essay, "statistical power" actually means something slightly different from this.

can get absolute change $C_j = B_j - F_j$ and percentage change $P_j = (B_j - F_j)/B_j$ for patient j by calculating from the baseline B_j and follow-up F_j scores immediately.

In this example, j is the patients' ID number, and here $n = 5$. In columns 2 and 3, there are baseline and follow-up scores for each patient. Absolute change and percentage change that calculated from baseline and follow-up scores are shown in column 4 and 5, respectively.

Table 1. Supine systolic blood pressure (in mmHg) for 5 patients with moderate essential hypertension, immediately before and after taking the drug captopril³

ID (j)	Baseline (B_j)	Follow-up (F_j)	Absolute (C_j)	Percentage(%) (P_j)
1	210	201	9	4.3
2	169	165	4	2.4
3	187	166	21	11.2
4	160	157	3	1.9
5	167	147	20	12.0

$$(C_j = B_j - F_j, P_j = 100 \times (B_j - F_j)/B_j)$$

From Table 1, we see that there is a decreasing effect for the blood pressure of each patient after taking the drug captopril. Absolute change and percentage change show the decrease in different ways. From a statistical point of view, we should compare the statistical power for the two methods.

2.3 Comparison of Statistical Power

We have mentioned that Vickers (2001) compared the statistical power of different methods by doing a simulation. However, his conclusion just based on an ideal simulation procedure, and he did not compare the statistical power theoretically. Kaiser (1989) developed a test statistic which compared the maximum likelihood of the two methods. It has nothing to do with statistical power. But Kaiser (1989) gave an idea that it is easier to do comparison by constructing a ratio test statistic.

For comparison of the statistical power of the two methods, we construct a ratio test statistic by using the test statistic or p-value of the treatment effect test. Before that, we need to know the distributions of absolute change and percentage change. That is because, for different distributions of absolute change or percentage change, different test methods will be used. In order to construct a ratio test statistic, we should know the test statistics used in both numerator and denominator of the ratio test statistic.

2.3.1 When t-test is Suitable for both Absolute Change and Percentage Change

When the sample size n of the clinical experiment is large, according to the *Central Limit Theorem*, both the mean of

absolute change and percentage change are asymptotic normally distributed, i.e.

$$\begin{aligned} \bar{C} &\sim N(\mu_C, \sigma_C^2) \\ \bar{P} &\sim N(\mu_P, \sigma_P^2) \end{aligned}$$

where μ_C and σ_C are the mean and the standard deviation of \bar{C} , μ_P and σ_P are the mean and the standard deviation of \bar{P} , respectively.

In this case, t-test can be used for both absolute change and percentage change. For absolute change, the null hypothesis of t-test is $H_0 : \mu_C = 0$. From $\bar{C} \sim N(\mu_C, \sigma_C^2)$, we get the t-statistic for an absolute change t-test is

$$t_C = \frac{\bar{C} - 0}{\hat{\sigma}_C} = \frac{\bar{C}}{\hat{\sigma}_C} \quad (3)$$

where $\hat{\sigma}_C$ is an estimate of the standard deviation of absolute change.

Similarly, in the percentage case, the null hypothesis of t-test is $H_0 : \mu_P = 0$. From $\bar{P} \sim N(\mu_P, \sigma_P^2)$, we get the t-statistic for a percentage change t-test is

$$t_P = \frac{\bar{P} - 0}{\hat{\sigma}_P} = \frac{\bar{P}}{\hat{\sigma}_P} \quad (4)$$

where $\hat{\sigma}_P$ is an estimate of the standard deviation of percentage change.

We have mentioned the relation between statistical power and the absolute value of t-statistic. We know that, when the significant level α is fixed, if the expected absolute value of the t-statistic of absolute change is larger, the statistical power of that will be higher. The opposite is also true, i.e.

$$\begin{aligned} R &= \frac{E(|t_C|)}{E(|t_P|)} = \frac{E\left(\left|\frac{\bar{C}}{\hat{\sigma}_C}\right|\right)}{E\left(\left|\frac{\bar{P}}{\hat{\sigma}_P}\right|\right)} > 1 \\ &\Leftrightarrow \frac{\text{Statistical Power of Absolute Change}}{\text{Statistical Power of Percentage Change}} > 1 \end{aligned} \quad (5)$$

where $E(|t_C|)$ and $E(|t_P|)$ are the expected absolute value of the t-statistic of absolute change and percentage change, respectively.

So, when $R > 1$, absolute change has higher statistical power than percentage change, and we choose absolute change. If $R < 1$, the percentage change with the higher statistical power is preferred.

In the case of small sample size, it is common to assume that one of the distributions of absolute change and percentage change is normal. In some specific situation, both absolute change and percentage change may be normally distributed. Even though for a dataset that is not normally distributed, if the distribution is close to normal distribution or

³Hand, DJ, Daly, F, Lunn, AD, McConway, KJ and Ostrowski, E (1994): A Handbook of Small Data Sets. London: Chapman and Hall. Dataset 72

the distribution is symmetric without extreme observations, t-test may also be used. In that situation, the test statistic R is also applicable.

If we simulate some datasets, by using the ratio test statistic R , we can compare the statistical power of the absolute change and percentage change of the datasets that we simulated. In contrast with Vickers' (2001) claim, some datasets with $R < 1$ will be shown, which reflects that percentage change has higher statistical power than absolute change under some conditions. In the simulation section, we will talk more about the comparison of the statistical power of the two methods.

2.3.2 When the t-test is not suitable for At Least One Test

For the cases when the assumptions for the t-test are not satisfied for at least one of the tests, another test should be considered. Wilcoxon rank sum test⁴ is an alternative method proposed by Wilcoxon (1945). Bonate (2000) mentioned that it is the non-parametric counterpart to the paired samples t-test and should be used when normal assumptions are violated. He suggested that the Wilcoxon rank sum test is always a better choice when the distribution of the data is unknown or uncertain. Therefore, when the paired samples t-test does not work, we choose Wilcoxon rank sum test instead.

Since we can not use t-statistic to construct the ratio test statistic any more, we may choose to use the expected p-value of the treatment effect test.

Similarly to (5), according to what is mentioned in equation (2), we may construct another ratio test statistic R' by taking the ratio of expected p-value, i.e.

$$R' = \frac{E(p_C)}{E(p_P)} < 1 \quad (6)$$

$$\Leftrightarrow \frac{\text{Statistical Power of Absolute Change}}{\text{Statistical Power of Percentage Change}} > 1$$

where $E(p_C)$ and $E(p_P)$ are the expected p-value of absolute change and percentage change, respectively.

For the cases that t-test still works, $E(p_C) = E(p_{t_C})$ and $E(p_P) = E(p_{t_P})$, where $E(p_{t_C})$ and $E(p_{t_P})$ are the expected p-value of the t-test for absolute change and percentage change, respectively. When Wilcoxon rank sum test is used instead of t-test, $E(p_C) = E(p_{C-Wilcoxon})$ and $E(p_P) = E(p_{P-Wilcoxon})$. $E(p_{C-Wilcoxon})$ and $E(p_{P-Wilcoxon})$ are the expected p-value of the Wilcoxon rank sum test for absolute change and percentage change, respectively. Therefore, we may get three different alternative forms for the ratio test statistic R' .

We have talked about that the smaller the expected p-value is, the higher the statistical power will be. When the significant level α is fixed, if $R' < 1$, absolute change has higher

statistical power than percentage change, and we choose absolute change. For $R' > 1$, the percentage change is preferred.

In this section, another ratio test statistic R' for non-normal distribution situation is discussed. This is the supplement of the normal distribution case. In the following simulation part, we will concentrate more on the normal distribution case shown in subsection 2.3.1, and the details in subsection 2.3.2 will not be discussed any more.

3. Rule of Thumb for the Standard Deviation of Change Scores

THE standard deviation of the treatment effect is an important parameter that is of interest in the planning of studies. The standard deviation of the change scores is the focus in the second section of this essay. For the case when t-test may be used instead of a non-parametric test, in order to calculate the ratio test statistic R , we should work out both the mean and the standard deviation of absolute change and percentage change first. It is easy to get these values in case the datasets of the experiment which give the baseline and follow-up scores for each patient are known. However, in clinical research, it is not always possible and practical to get the scores for each patient, especially in the planning phase of a study. If we require some datasets to support our research work, we may find some experiment datasets interesting for our research from experiments that someone else has done.

One of the good ways to find the datasets is searching from published clinical articles. Most of the time, we may find some examples in these articles which show us the summary of the baseline scores, the follow-up scores, and their standard deviations. And we may get relevant datasets from these tables. However, the scores for each patient in these clinical research articles are seldom published. In this case, how can we know the standard deviation of the change scores?

There is a rule of thumb which describes the relationship between the standard deviation of the change scores and that of the baseline scores.

$$SD(C) \approx \frac{SD(B)}{\sqrt{2}} \quad (7)$$

3.1 Theoretical Derivation of the Rule of Thumb for Absolute Change

The general expression for the rule of thumb of absolute change is

$$SD(C) = \frac{SD(B)}{k} \quad (8)$$

⁴Wilcoxon rank sum test is a non-parametric test for assessing whether two independent samples of observations come from the same distribution.

where k is a constant that should be determined; $SD(C)$ and $SD(B)$ are the standard deviation of absolute change and baseline scores, respectively.

The relationship between the standard deviation of absolute change scores and that of the baseline scores can be derived from properties of the variance of C ,

$$\begin{aligned} Var(C) &= Var(B - F) \\ &= Var(B) + Var(F) - 2cov(B, F) \\ &= Var(B) + Var(F) - 2r\sqrt{Var(B)Var(F)} \end{aligned} \quad (9)$$

where $r = cov(B, F) / \sqrt{Var(B)Var(F)}$ is the correlation coefficient between baseline and follow-up scores.

We assume that $Var(F) = mVar(B)$, where m is the ratio of the variance of follow-up scores to that of baseline scores. In a specific case, m is a constant which may be calculated from the known dataset. Then, equation (9) can be rewritten as

$$\begin{aligned} Var(C) &= Var(B) + Var(F) - 2r\sqrt{Var(B)Var(F)} \\ &= Var(B) + mVar(B) - 2r\sqrt{Var(B)mVar(B)} \\ &= (1 + m)Var(B) - 2r\sqrt{m}Var(B) \\ &= (1 + m - 2r\sqrt{m})Var(B) \end{aligned} \quad (10)$$

Equivalently, from equation (10), we get the relation equation (11).

$$SD(C) = \sqrt{1 + m - 2r\sqrt{m}}SD(B) \quad (11)$$

Thus, in equation (8), $k = 1/\sqrt{1 + m - 2r\sqrt{m}}$, which shows that the rule of thumb is determined by the correlation coefficient r and the ratio m .

When the baseline and follow-up scores have the same variance, $m = 1$, then we get $k = 1/\sqrt{2 - 2r}$. And the expression of the rule of thumb becomes

$$SD(C) = \sqrt{2 - 2r}SD(B) \quad (12)$$

Therefore, the standard deviation of the absolute change is connected to the standard deviation of the baseline scores via the correlation coefficient r .

3.2 The Relation between the Standard Deviation of Absolute Change and the Correlation Coefficient

The empirical form of the rule of thumb is shown in expression (7). From $k = 1/\sqrt{2 - 2r} \approx \sqrt{2}$, we get $r \approx 0.75$. Equation (12) shows the relation between the standard deviation of the absolute change and that of the baseline scores. So, how does the standard deviation of absolute change depend on the correlation coefficient?

If we assume that $SD(B) = 1$, then we get $SD(C) = \sqrt{2 - 2r}$. The smooth curve in Figure 1 shows the relationship between the standard deviation of the absolute change $SD(C)$ and the correlation coefficient r . We see that $SD(C)$ decreases from 1.4 to 0 as the correlation coefficient increases from 0 to 1. When $r \leq 0.8$, $SD(C)$ roughly has a linear decrease. After that, when the correlation coefficient r tends to 1, the ratio decreases quickly to 0.

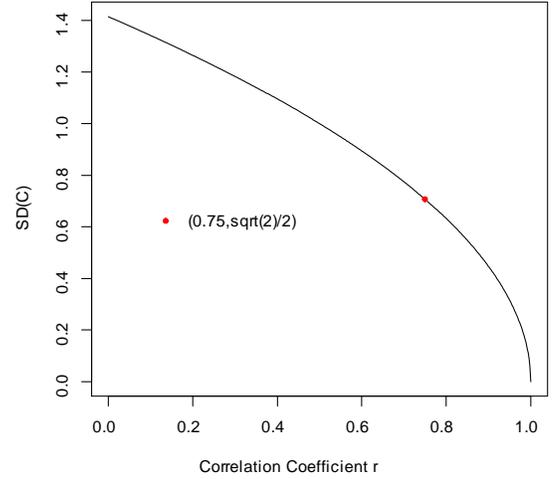


Figure 1. Relation curve between the standard deviation of absolute change $SD(C)$ and the correlation coefficient r when $SD(B) = 1$.

Therefore, when $r \approx 0.75$, the empirical form of the rule of thumb $SD(C) \approx SD(B)/\sqrt{2}$ holds. If the correlation coefficient r changes to another value, the form of the rule of thumb will be also changed. When r tends to 1, a little change in r may result in a significant change in the standard deviation of absolute change.

3.3 Rule of Thumb for Percentage Change

Earlier we stated that $P = (B - F)/B = C/B$, i.e. percentage change is the ratio of absolute change to the baseline score. Then we get

$$SD(P) = SD\left(\frac{B - F}{B}\right) = SD\left(\frac{C}{B}\right)$$

Since percentage change is a ratio of two variables, its distribution is uncertain. It is hard to derive the expression of the standard deviation of percentage change from the standard deviation of baseline scores as we did in equation (10).

We have discussed the rule of thumb for $SD(C)$, and we see that the standard deviation of percentage change depends

not only on the absolute change C but also on the baseline score B . If we fix the value of C , B may also keep on changing from one sample to another. As a result, it seems there is no stable relationship between the standard deviation of percentage change $SD(P)$ and the baseline score B . A rule of thumb for percentage change may, therefore, not be stated. This conclusion will be proved in the following simulation part.

4. Simulation

IN section 2, we discussed the comparison of the statistical power of absolute change and percentage change, by constructing a ratio test statistic based on normal distribution. In the third section, we discussed the rule of thumb for absolute change and percentage change theoretically.

This section will do some simulations to show the problems that we have discussed in a practical way. The first thing we want to prove is, in contrast with Vickers' (2001) conclusion, that percentage change can be statistically efficient under some conditions. The second thing that will be proved is the difficulty of defining a rule of thumb for percentage change.

4.1 Statistical Efficiency of Percentage Change under Some Conditions

Vickers (2001) suggested avoiding using percentage change, because of his conclusion that percentage change from baseline is statistically inefficient. He made that conclusion based on the comparison of statistical power calculated from his simulation results.

Vickers (2001) did the simulation in the following way. First, he simulated 100 pairs of baseline and follow up scores for 100 patients. The baseline scores B are simulated from a normal distribution, i.e. $B \sim N(50, 10)$. In order to get 100 scores B , he simulated 100 B' first, $B' \sim N(0, 10)$, then he got B from the equation $B = B' + 50$. He also simulated another 100 scores Y , $Y \sim N(0, 10)$, which are defined as the post-treatment scores of the control group. Then the follow-up scores F' are simulated from B' and Y by using the equation (13). We should note that F' is not the final follow-up scores.

$$F' = B'r + Y \cdot \sqrt{1 - r^2} + 50 \quad (13)$$

From $B' \sim N(0, 10)$ and $Y \sim N(0, 10)$, we obtain that $F \sim N(50, 10)$. Finally, Vickers (2001) simulated 100 g from *Binomial*(1, 0.5) for each patient. These patients who got $g = 1$ were put into the treatment group, and the other patients were put in the control group. So, there are nearly 50 patients in both treatment group and control group. For

the patients in the treatment group, the final follow-up scores F all have an absolute decrease of 5 units from F' after the medical intervention, while there is no change of the follow-up scores for patients in the control group, i.e.

$$F = \begin{cases} F' - 5 & \text{if } g = 1 \\ F' & \text{if } g = 0 \end{cases}$$

He changed the correlation coefficient r , and got different simulation results under different correlation coefficient. Using these simulation results, he calculated statistical power for each method and made the statistical inefficiency conclusion.

In the following subsections, we will do simulations based on Vickers' (2001) method. But some change and improvement will be done to his code.

4.1.1 A Case that Percentage Change Has Higher Statistical Power

In Vickers' (2001) simulation method, a fixed absolute change from the simulated follow-up scores F' to the final follow-up scores F was set to each patient in the treatment group. If we change the fixed absolute change to a fixed percentage change, maybe we will get something different. To be more randomized, just like what may happen in practice, we use a random percentage change instead of fixed percentage change. The percentage changes P are simulated from a normal distribution. We should notice that the changes we did to Vickers' (2001) simulation will result in the change of the correlation coefficient between the baseline and follow-up scores. The correlation coefficient of the baseline and follow-up scores in the simulation result is not the r that we used in equation (13) any more, even though the real value of the correlation coefficient may be very close to the value we used in the simulation. However, since the correlation coefficient will not affect the comparison of the statistical power for the two methods, we will give the value of used in each simulation procedure, but do not talk more about it. What's more, in the following simulation, we just concentrate on the patients in the treatment group.

We have developed a ratio test statistic R in section 2, and we will use it to do the comparison between absolute change and percentage change. The simulation can be divided into two steps. In step 1, we simulate 100 pairs of baseline/follow-up scores. In the second step, the test statistic R is calculated based on the scores we simulated in step 1.

From equation (5), we know that, in order to simulate a dataset such that $R < 1$, we should let the percentage change have a large mean and small standard deviation. So, in this case, we simulate P from the normal distribution $N(0.5, 0.01)$. We set $r = 0.75$ and simulate B from the distribution $N(200, 20)$. According to Vickers (2001) simulation method, we obtain a dataset of scores. Figure 2 shows a part of the simulation results of baseline and follow-up scores in

step 1. From Figure 2, we see that there is a nearly 50% decrease from the baseline score for each patient.

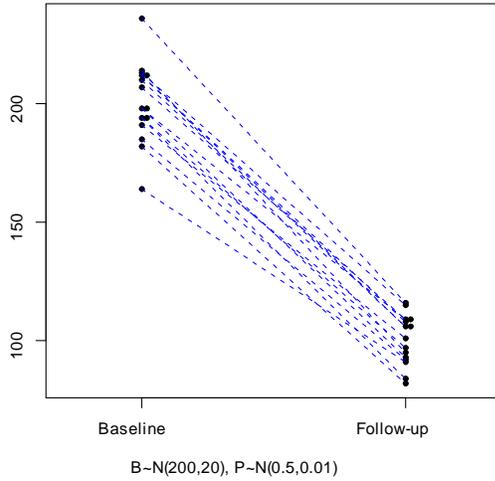


Figure 2. Change from Baseline Scores to Follow-up Scores ($r = 0.75$).

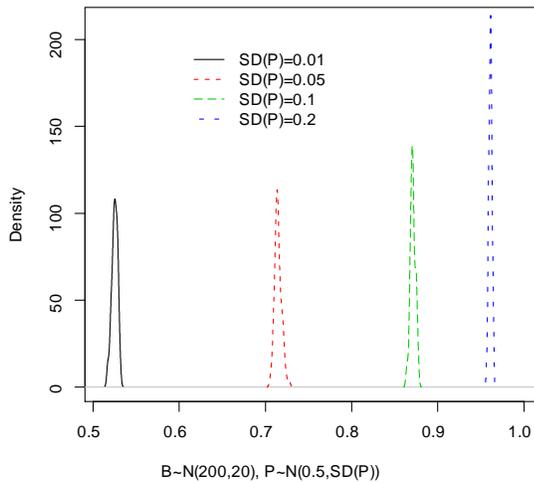


Figure 3. Distribution of R ($r = 0.75$).

Since the test statistic R is the ratio of two expected values, in order to calculate the expected value, we repeat the score simulation procedure in step 1 100 times, and then we get 100 datasets of scores. In the second step, using the 100 datasets, we work out the value of R , and check if it is less than 1.

In order to show a more general result, we repeat the procedure in both step 1 and step 2 100 times, and check the distribution of R . As shown in Figure 3, the solid line on the left is the distribution of R based on the datasets we simulated. We see that, when $P \sim N(0.5, 0.01)$, the value of R is much less than 1. As a result, in this case, percentage change has a higher statistical power.

In this simulation, we set the percentage change P normally distributed with a large mean and small standard deviation. However, it is unreasonable to have such a small standard deviation in practice. If we increase the standard deviation of P , what kind of result will come to us?

Figure 3 shows the distribution of R under different standard deviation of P . We see that, the value of the test statistic R increases as the standard deviation of P increases. Although R increases, it is still less than 1. In this case, we prefer percentage change to absolute change.

4.1.2 A Case that Absolute Change Has A Little Higher Statistical Power

In the last section, we simulated a case where percentage change had a higher statistical power, which is in contrast with Vickers' (2001) conclusion. If we consider more about the simulation method, we should notice that we used percentage change to do simulation in that case. It may be a factor which affects the simulation results such that percentage change has a higher statistical power.

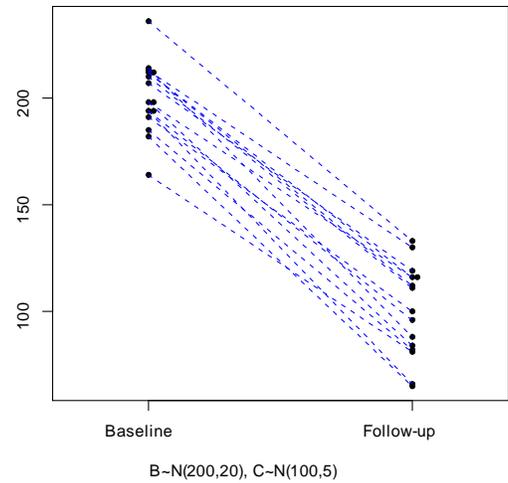


Figure 4. Change from Baseline Scores to Follow-up Scores ($r = 0.75$).

Now, we just change $P \sim N(0.5, 0.01)$ to $C \sim N(100, 5)$, and keep other conditions the same. Part of the baseline and

follow-up scores are shown in Figure 4. It seems similar with the scores in Figure 2. This is because we set the expected absolute change to 100, which is 50% of baseline scores. So, in a similar absolute change case, what kind of result we will get?

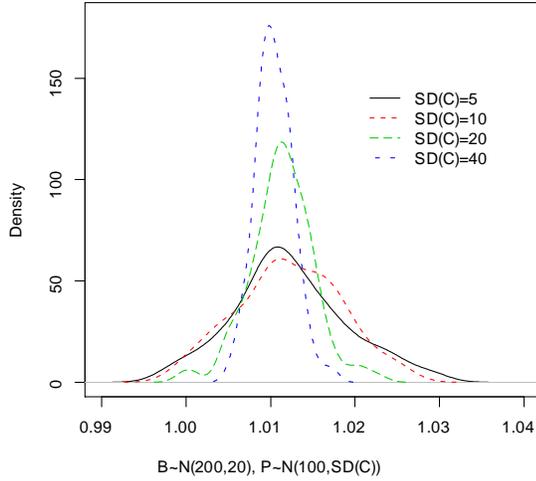


Figure 5. Distribution of R ($r = 0.75$).

Figure 5 shows the distribution of R under different standard deviation of C . Comparing with the distributions in Figure 3, we find it has a different kind of change when the standard deviation of C changes. The distributions in Figure 3 mainly perform a location difference, while the distributions in Figure 5 have different kurtosis and spread.

Even though the expected value of R in Figure 5 is larger than 1, it is really close to 1. In this case, it seems both absolute change and percentage change can be used. The difference between the statistical powers of the two methods is very small.

4.1.3 Another Case that Percentage Change Has Little Difference with Absolute Change

In this case, we reduce the standard deviation of the baseline scores to 10 and compare the results with that of the previous case.

From Figure 6, we find that the expected value of the ratio test statistic is much more close to 1. If we also reduce the mean of C , then R is completely less than 1. This is another case that shows, under some conditions, the statistical powers of the two methods are nearly the same.

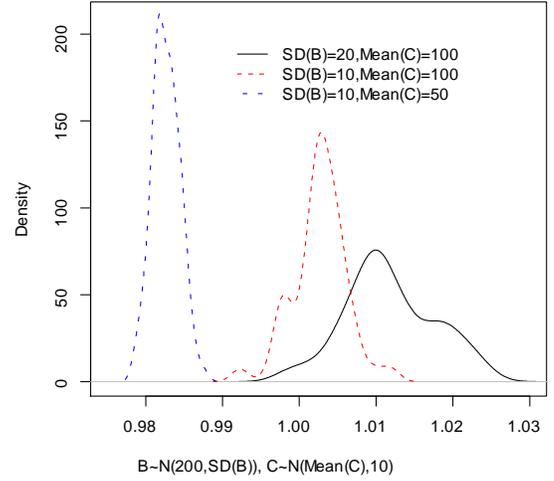


Figure 6. Distribution of R ($r = 0.75$).

We have done 3 simulations based on a modification of Vickers' (2001) method so far. The first one shows that percentage change has higher statistical power. The other two show that percentage has nearly the same statistical power with absolute change. All of them proved that percentage change can be statistical efficient under some conditions. Therefore, Vickers (2001) conclusion is not correct.

4.2 Nonexistence of Rule of Thumb for Percentage Change

We have discussed the rule of thumb for percentage change theoretically in section 3. In this section we will simulate another dataset to check if the rule of thumb for percentage change exists. The simulation will show how the standard deviation of percentage change $SD(P)$ depends on the baseline scores B .

This simulation is also based on Vickers' (2001) simulation method. In this case, the baseline scores follows $B \sim N(50, 10)$, and the percentage change has a distribution $P \sim N(0.1, 0.02)$. Following the simulation steps, we will get a score dataset, and the standard deviations of absolute change and percentage change can be calculated.

After we get the baseline and follow-up scores, we make a simple transformation that both baseline and follow-up scores decrease 5 units, i.e.

$$\begin{aligned} \tilde{B} &= B - 5 \\ \tilde{F} &= F - 5 \end{aligned}$$

After transformation, we get a new dataset of baseline and follow-up scores, and calculate the standard deviations of absolute change and percentage change of new scores. Repeat

the simulation procedure 100 times, each time we may get 4 standard deviations, $SD(C)$, $SD(\tilde{C})$, $SD(P)$, $SD(\tilde{P})$. Then, we calculate the mean of 100 simulation results for the 4 standard deviations, respectively. When the correlation coefficient changes, we get the relation curve between the standard deviation of change scores and the correlation coefficient both before and after transformation.

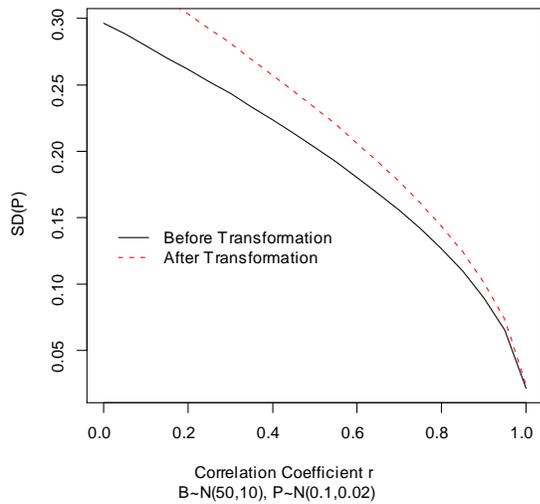
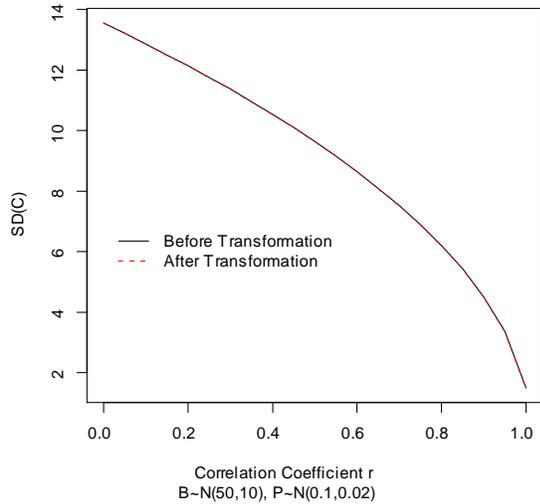


Figure 7. Relation curve between the standard deviation of absolute change (above) or percentage change (below) and the correlation coefficient r before and after transformation.

From figure 7, we observe that, after transformation, the

standard deviation of absolute change does not change. Actually, we can prove that in a theoretical way.

$$\tilde{C} = \tilde{B} - \tilde{F} = (B - 5) - (F - 5) = B - F = C$$

After the transformation, C does not change. Therefore, the standard deviation of absolute change will not change, neither.

From figure 7, we also see that, the standard deviations of percentage change under different correlation coefficients become larger after transformation. The smaller the correlation coefficient is, the larger the change of the standard deviation of percentage change will be.

We have mentioned that $SD(P) = SD(C/B)$. In this case, C does not change, but B becomes smaller. As a result, the standard deviation of percentage change becomes larger. This reflects that the standard deviation of percentage change depends on the baseline scores. Therefore, it's difficult to have a rule of thumb for the standard deviation of the percentage change based on only the standard deviation of the baseline scores.

5. Demonstration of the Rule of Thumb for Absolute Change

WE have discussed the rule of thumb for absolute change theoretically in the third section of this essay. A general expression of the rule of thumb is given in equation (12). We know that, when $r \approx 0.75$, the empirical form of the rule of thumb in expression (7) holds. In this section, we will concentrate on the demonstration of the rule of thumb for absolute change. We collect some real datasets to check whether the rule of thumb works well.

The real datasets are searched from examples of clinical research articles and medical literatures. As shown in Table 2, we have collected two kinds of datasets. The datasets of the first 10 cases contain the scores for each patient, while the data sets of the last 5 cases just contain some data summary, e.g. mean, standard deviation, etc., of the baseline scores and absolute change scores.

If we know the scores for each patient, we can calculate not only the standard deviations but also the correlation coefficient between the baseline scores and the follow-up scores. So, for the first 10 cases, we also get the value for the general form of the rule of thumb.

Table 2. Comparison of the real standard deviation and the value got from the rule of thumb for absolute change

Case	$SD(B)$	$SD(F)$	$SD(C)$	$SD(B)/\sqrt{2}$	$\sqrt{2-2r}SD(B)$	r	m
1 ⁵	11.43	8.43	8.99	8.09	9.86	0.63	0.54
2 ⁶	12.29	6.94	7.91	8.69	7.76	0.8	0.32
3 ⁷	5.59	5.14	2.54	3.95	2.61	0.89	0.84
4 ⁸	4.79	4.86	2.68	3.39	2.66	0.85	1.03
5 ⁹	6.32	13.32	13.76	4.47	8.15	0.17	4.45
6 ¹⁰	20.57	20	9.03	14.54	9.14	0.9	0.95
7 ¹¹	13.19	19.02	17.26	9.33	13.54	0.47	2.08
8 ¹²	15.61	18.3	7.98	11.04	6.94	0.9	1.37
9 ¹³	21.85	24.13	27.05	15.45	25.65	0.31	1.22
10 ¹⁴	4.27	4.39	3.78	3.02	3.72	0.62	1.06
11 ¹⁵	0.93		0.37	0.66			
12 ¹⁶	12.5		3.8	8.84			
13 ¹⁷	18.3		6.3	12.94			
14 ¹⁸	18		10	12.73			
15 ¹⁹	16		15	11.31			

Comparing the values of $SD(C)$ and $SD(B)/\sqrt{2}$, there are obvious differences between the two values for these real datasets. In some cases, the difference between $SD(C)$ and $SD(B)/\sqrt{2}$ is very large.

For the cases that has a correlation coefficient between 0.6 and 0.9, which is close to 0.75, the difference between $SD(C)$ and $SD(B)/\sqrt{2}$ may be acceptable. For example, in case 1, $r = 0.63$, it is close to 0.75. In this case, $SD(C) = 8.99$ and $SD(B)/\sqrt{2} = 8.09$, it seems that the two values are close to each other. However, in case 5, when $r = 0.17$, the value of $SD(C)$ is nearly three times of the value of $SD(B)/\sqrt{2}$. This is not acceptable. These facts show that the rule of thumb in expression (7) is valid when $r \approx 0.75$ or when r is close to that value.

If we take the correlation coefficient r into account, by comparing the values of $SD(C)$ and $\sqrt{2-2r}SD(B)$, we find that the values of $\sqrt{2-2r}SD(B)$ are closer to $SD(C)$ than $SD(B)/\sqrt{2}$, especially when $m \approx 1$. Look at case

4, $m = 1.03 \approx 1$. In this case, $SD(C) = 2.68$ and $\sqrt{2-2r}SD(B) = 2.66$, the two values are nearly the same, which reflect that the rule of thumb are also affected by the ratio $m = Var(F)/Var(B)$.

If we also take m into account, we will get the real value of the standard deviation of absolute change. Actually, we have proved that in equation (11) of section 3.

When we know nothing about the correlation between the baseline and follow-up scores, just like the last five cases in Table 2, the rule of thumb may not be suitable. In this case, we should be more careful.

From the analysis based on real datasets in this section, we learned that when the ratio $m = Var(F)/Var(B)$ tends to 1 and the correlation coefficient is nearly 0.75, the rule of thumb $SD(C) \approx SD(B)/\sqrt{2}$ will be practical. If these conditions are not satisfied, it is not a good rule to follow. If we ignore the two conditions and insist on using the rule, as we know from Table 2, it may result in a big mistake.

⁵Douglas G.Alman (1991): Practical Statistics for Medical Research. London: Chapman and Hall. Page 475
⁶Douglas G.Alman (1991): Practical Statistics for Medical Research. London: Chapman and Hall. Page 475
⁷Pagano M, Gauvreau K (2000): Principles of Biostatistics, Second Edition, Duxbury. Table B.15
⁸Pagano M, Gauvreau K (2000): Principles of Biostatistics, Second Edition, Duxbury. Table B.15
⁹Bradstreet, T.E. (1994) "Favorite Data Sets from Early Phases of Drug Research - Part 3." Proceedings of the Section on Statistical Education of the American Statistical Association. <<http://www.math.iup.edu/~tshort/Bradstreet/part3/part3-table3.html>> 2009-06-08
¹⁰Hand, DJ, Daly, F, Lunn, AD, McConway, KJ and Ostrowski, E (1994): A Handbook of Small Data Sets. London: Chapman and Hall. Dataset 72
¹¹Ryan, Joiner, Cryer (1985): Minitab Handbook, Second Edition. PWS-KENT Publishing Company. Page 318, Pulse Data
¹²Bonate P (2000): Analysis of Pretest-Posttest Design. Boca Raton: Chapman and Hall/CRC. Table 3.1
¹³Bonate P (2000): Analysis of Pretest-Posttest Design. Boca Raton: Chapman and Hall/CRC. Table 3.4
¹⁴Bonate P (2000): Analysis of Pretest-Posttest Design. Boca Raton: Chapman and Hall/CRC. Table 9.1
¹⁵Waleekhachonloet O, Limwattananon C, Limwattananon S, Gross C (2007): Group behavior therapy versus individual behavior therapy for healthy dieting and weight control management in overweight and obese women living in rural community. Obesity Research & Clinical Practice, 1: 223-232. Table 3
¹⁶Neovius M, Rössner S (2007): Results from a randomized controlled trial comparing two low-calorie diet formulae. Obesity Research & Clinical Practice, 1: 165-171. Table 1 & Table 2
¹⁷Neovius M, Rössner S (2007): Results from a randomized controlled trial comparing two low-calorie diet formulae. Obesity Research & Clinical Practice, 1: 165-171. Table 1 & Table 2
¹⁸Neovius M, Rössner S (2007): Results from a randomized controlled trial comparing two low-calorie diet formulae. Obesity Research & Clinical Practice, 1: 165-171. Table 1 & Table 2
¹⁹Neovius M, Rössner S (2007): Results from a randomized controlled trial comparing two low-calorie diet formulae. Obesity Research & Clinical Practice, 1: 165-171. Table 1 & Table 2

6. Discussion and Conclusions

IN this essay we compared the use of absolute change and percentage change. According to the definition of statistical power, we developed a ratio test statistic R under certain distribution assumption, which can help us decide which method will be used, absolute change or percentage change. When $R > 1$, absolute change has a higher statistical power. In that case, we prefer absolute change to percentage change. If $R < 1$, we choose percentage change.

Based on Vickers' (2001) simulation method, with the help of the ratio test statistic, we did some simulations to compare the statistical power of the two methods. In contrast with Vickers' (2001) conclusion that the percentage change is statistical inefficient, we simulated some datasets in which percentage change has higher statistical power, or has nearly the same statistical power with absolute change. In this way, we showed that percentage can be statistical efficient under some conditions.

Another issue often concerned by researchers is the standard deviation of change scores. There is a rule of thumb that may help us get the standard deviation of change scores quickly from the standard deviation of the baseline scores. The general form of the rule of thumb for absolute change can be derived in a theoretical way. From the derivation in section 3, we know that, when the ratio $m = Var(F)/Var(B) \approx 1$ and $r \approx 0.75$, the empirical form of the rule of thumb $SD(C) \approx SD(B)/\sqrt{2}$ holds. We also checked these conditions in a practical way by collecting some real data to compare the real value and the value got from the rule of thumb. And we got the same conclusion. So, if we know nothing about the correlation between baseline and follow-up scores, we should be more careful to use the rule.

For percentage change, a rule of thumb does not exist. That is because the standard deviation of percentage change depends on the baseline scores, and it is very hazardous to state a rule. We also proved this by doing a simulation with a simple transformation. The simulation result showed how the standard deviation of percentage change depends on the baseline scores.

In this essay, we didn't give any rules to make a choice between absolute change and percentage change. We developed a ratio test statistic which may be helpful, but it is not a good rule to tell us how to make a choice. That is because, in the ratio test statistic, expected value should be used. For a specific dataset in practice, we can not calculate the expected value. That is the limitation of the test statistic.

Actually, there is not a most optimal method to tell us which method to choose. From a statistical point of view, we would like to choose the method with higher statistical power. Beside the two change measurement methods from baseline, there are also some other methods. One of them is analysis of covariance, which is mentioned by both Vickers (2001) and Kaiser (1989). Bonate (2000) discussed this Zhang, L. and Han, K. (2009)

in more detail. It has higher statistical power than the two methods we talked about. But for the people who are not statisticians, this method can not be understood as easily as absolute change or percentage change.

From a clinical point of view, clinicians may prefer to choose the method that will show the health-improvement more obviously. Some researchers may choose the method that may be understood by most people that are interested in his research. Sometimes, we may make a choice just based on some empirical information. Most of the time, the choice depends more on the research work and the researcher's own experience.

References

- BONATE P. (2000): *Analysis of Pretest-Posttest Design*. Boca Raton: Chapman and Hall/CRC.
- KAISER L. (1989): Adjusting for baseline: change or percentage change? *Statistics in Medicine*, 10: 1183-1190.
- KIM M. et al. (2009): Comparison of epicardial, abdominal and regional fat compartments in response to weight loss. *Nutr Metab Cardiovasc Dis*, 1:7. doi: 10.1016/j.numecd.2009.01.010.
- LAVANGE L., ENGELS J. and ACCURSO F. (2007): Analyzing percent change in cystic fibrosis clinical trials. The 21st Annual North American Cystic Fibrosis Conference, Anaheim, California.
- NEOVIUS M. and RÖSSNER S. (2007): Results from a randomized controlled trial comparing two low-calorie diet formulae. *Obesity Research & Clinical Practice*, 1: 165-171.
- TÖRNQVIST L., VARTIA P. and VARTIA Y. (1985): How should relative changes be measured? *American Statistician*, 39: 43-46.
- VICKERS A. (2001): The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology*, 1:6.
- WALEEKHACHONLOET O., LIMWATTANANON C., LIMWATTANANON S. and GROSS C. (2007): Group behavior therapy versus individual behavior therapy for healthy dieting and weight control management in overweight and obese women living in rural community. *Obesity Research & Clinical Practice*, 1: 223-232.
- WILCOXON F. (1945): Individual comparisons by ranking methods, *Biometrics Bull.*, 1, 80.

Appendix

*R Code*²⁰

Figure 1

```
r<-SDC<-NULL
for(i in 1:501){
  r[i]<-0.002*(i-1)
  SDC[i]<-sqrt(2*(1-r[i]))
}
plot(r,SDC,xlab="Correlation Coefficient r",ylab="SD(C)",type="l")
points(0.75,sqrt(2*(1-0.75)),pch=20,col=2)
legend(0.1,0.7,"(0.75,sqrt(2)/2)",pch=20,col=2,bty="n")
```

Figure 2

```
rm(list=ls())
set.seed(12345)
n<-15
mu<-0
sd<-20
b<-rnorm(n,mu,sd)
y<-rnorm(n,mu,sd)
r<-0.75 #correlation coefficient
f<-b*r+y*(1-r^2)^0.5+200
h<-rnorm(n,0.5,0.01) #percentage change
f<-f-(f*h)
f<-round(f) #follow-up score
b<-round(b)+200 #baseline score
fun<-function(b,f){
  l<-list(b,f)
  stripchart(l,vertical=T,group.names=c("Baseline","Follow-up"),xlim=c(0.7,2.3),pch=20,
    method="stack",main="Change from Baseline to Follow-up")
  for(i in (1:length(b))){
    lines(c(1,2),c(b[i],f[i]),lty=3,col=4)
  }
  mtext(side=1,line=3,"B~N(200,20), P~N(0.5,0.01)")
}
fun(b,f)
```

Figure 3

```
rm(list=ls())
set.seed(12345)
n<-100
mu<-0
sd<-20
R<-NULL
mc<-mp<-sdc<-sdp<-atc<-atp<-NULL
for(s in c(0.01,0.05,0.1,0.2)){
```

²⁰Responsible Programmer: Ling Zhang.

```

for(k in 1:100){
  for(j in 1:100){
    b<-rnorm(n,mu,sd)
    y<-rnorm(n,mu,sd)
    r<-0.75 #correlation coefficient
    f<-b*r+y*(1-r^2)^0.5+200
    h<-rnorm(n,0.5,s) #percentage change
    f<-f-(f*h)
    f<-round(f) #follow-up score
    b<-round(b)+200 #baseline score
    c<-b-f
    p<-(b-f)/b
    mc<-mean(c)
    mp<-mean(p)
    sdc<-sd(c)
    sdp<-sd(p)
    atc[j]<-abs(mc/sdc) #absolute value of tc, i.e. |tc|
    atp[j]<-abs(mp/sdp) #absolute value of tc, i.e. |tp|
  }
  atc<-na.omit(atc)
  atp<-na.omit(atp)
  eatc<-mean(atc) #expected value of |tc|
  eatp<-mean(atp) #expected value of |tp|
  R[k]<-eatc/eatp #value of the ratio test statistic R
}
if(s<=0.01){
  plot(density(R),xlim=c(0.51,1),ylim=c(0,210),xlab="B~N(200,20), P~N(0.5,SD(P))",
    ylab="Density",main="Distribution of R")
}
else{
  if(s<=0.05){
    lines(density(R),col=2,lty=2)
  }
  else{
    if(s<=0.1){
      lines(density(R),col=3,lty=3)
    }
    else{
      lines(density(R),col=4,lty=4)
    }
  }
}
}
legend(0.6,200,c("SD(P)=0.01","SD(P)=0.05","SD(P)=0.1","SD(P)=0.2"),lty=c(1,2,3,4),
  col=c(1,2,3,4),bty="n")

```

Figure 4

```

rm(list=ls())
set.seed(12345)
n<-15
mu<-0
sd<-20
b<-rnorm(n,mu,sd)

```

```

y<-rnorm(n,mu,sd)
r<-0.75 #correlation coefficient
f<-b*r+y*(1-r^2)^0.5+200
h<-rnorm(n,100,5) #absolute change
f<-f-h
f<-round(f) #follow-up score
b<-round(b)+200 #baseline score
fun<-function(b,f){
  l<-list(b,f)
  stripchart(l,vertical=T,group.names=c("Baseline","Follow-up"),xlim=c(0.7,2.3),pch=20,
    method="stack",main="Change from Baseline to Follow-up")
  for(i in (1:length(b))){
    lines(c(1,2),c(b[i],f[i]),lty=3,col=4)
  }
  mtext(side=1,line=3, "B~N(200,20), C~N(100,5)")
}
fun(b,f)

```

Figure 5

```

rm(list=ls())
set.seed(12345)
n<-100
mu<-0
sd<-20
R<-NULL
mc<-mp<-sdc<-sdp<-atc<-atp<-NULL
for(s in c(5,10,20,40)){
  for(k in 1:100){
    for(j in 1:100){
      b<-rnorm(n,mu,sd)
      y<-rnorm(n,mu,sd)
      r<-0.75 #correlation coefficient
      f<-b*r+y*(1-r^2)^0.5+200
      h<-rnorm(n,100,s) #absolute change
      f<-f-h
      f<-round(f) #follow-up score
      b<-round(b)+200 #baseline score
      c<-b-f
      p<-(b-f)/b
      mc<-mean(c)
      mp<-mean(p)
      sdc<-sd(c)
      sdp<-sd(p)
      atc[j]<-abs(mc/sdc) #absolute value of tc, i.e. |tc|
      atp[j]<-abs(mp/sdp) #absolute value of tp, i.e. |tp|
    }
    atc<-na.omit(atc)
    atp<-na.omit(atp)
    eatc<-mean(atc) #expected value of |tc|
    eatp<-mean(atp) #expected value of |tp|
    R[k]<-eatc/eatp #value of the ratio test statistic R
  }
}
if(s<=5){

```

```

plot(density(R),xlim=c(0.99,1.04),ylim=c(0,180),xlab="B~N(200,20), P~N(100,SD(C))",
     ylab="Density",main="Distribution of R")
}
else{
  if(s<=10){
    lines(density(R),col=2,lty=2)
  }
  else{
    if(s<=20){
      lines(density(R),col=3,lty=3)
    }
    else{
      lines(density(R),col=4,lty=4)
    }
  }
}
}
legend(1.02,150,c("SD(C)=5", "SD(C)=10", "SD(C)=20", "SD(C)=40"),lty=c(1,2,3,4),
      col=c(1,2,3,4),bty="n")

```

Figure 6

```

rm(list=ls())
set.seed(12345)
n<-100
mu<-0
R<-NULL
mc<-mp<-sdc<-sdp<-atc<-atp<-NULL
for(sd in c(20,10)){
  for(muc in c(100,50)){
    for(k in 1:100){
      for(j in 1:100){
        b<-rnorm(n,mu,sd)
        y<-rnorm(n,mu,sd)
        r<-0.75 #correlation coefficient
        f<-b*r+y*(1-r^2)^0.5+200
        h<-rnorm(n,muc,10) #absolute change
        f<-f-h
        f<-round(f) #follow-up score
        b<-round(b)+200 #baseline score
        c<-b-f
        p<-(b-f)/b
        mc<-mean(c)
        mp<-mean(p)
        sdc<-sd(c)
        sdp<-sd(p)
        atc[j]<-abs(mc/sdc) #absolute value of tc, i.e. |tc|
        atp[j]<-abs(mp/sdp) #absolute value of tp, i.e. |tp|
      }
      atc<-na.omit(atc)
      atp<-na.omit(atp)
      eatc<-mean(atc) #expected value of |tc|
      eatp<-mean(atp) #expected value of |tp|
      R[k]<-eatc/eatp #value of the ratio test statistic R
    }
  }
}

```

```

}
if(sd>=20&mu<=100){
  plot(density(R),xlim=c(0.975,1.03),ylim=c(0,210),
        xlab="B~N(200,SD(B)), C~N(Mean(C),10)",ylab="Density",main="Distribution of R")
}
else{
  if(sd>=10&mu<=100){
    lines(density(R),col=2,lty=2)
  }
  else{
    if(sd<=10){
      lines(density(R),col=4,lty=3)
    }
    else{
    }
  }
}
}
}
legend(0.99,200,c("SD(B)=20,Mean(C)=100","SD(B)=10,Mean(C)=100","SD(B)=10,Mean(C)=50"),
        lty=c(1,2,3),col=c(1,2,4),bty="n")

```

Figure 7

```

rm(list=ls())
par(mfrow=c(1,2),pty="s")
n<-100
mu<-0
sd<-10
SDC<-SDC2<-SDP<-SDP2<-matrix(0,100,21)
for(j in 1:100){
  b<-rnorm(n,mu,sd)
  y<-rnorm(n,mu,sd)
  r<-NULL
  bb<-b
  for(i in 1:21){
    r[i]<-0.05*(i-1) #correlation coefficient
    f<-b*r[i]+y*(1-r[i]^2)^0.5+50
    h<-rnorm(n,0.1,0.02) #percentage change
    f<-f-(f*h)
    f<-round(f) #follow-up score before transformation
    f2<-f-5 #follow-up score after transformation
    b<-round(b)+50 #baseline score before transformation
    b2<-b-5 #baseline score after transformation
    c<-b-f
    c2<-b2-f2
    p<-c/b
    p2<-c2/b2
    SDC[j,i]<-sd(c)
    SDC2[j,i]<-sd(c2)
    SDP[j,i]<-sd(p)
    SDP2[j,i]<-sd(p2)
    b<-bb
  }
}

```

```

}
MSDC<-MSDC2<-MSDP<-MSDP2<-NULL
for(i in 1:21){
  MSDC[i]<-mean(SDC[,i])
}
for(i in 1:21){
  MSDC2[i]<-mean(SDC2[,i])
}
plot(r,MSDC,type="l",xlab="Correlation Coefficient r",ylab="SD(C)")
lines(r,MSDC2,type="l",lty=2,col=2)
legend(0,7,c("Before Transformation",
  "After Transformation"),lty=c(1,2),col=c(1,2),bty="n")
mtext(side=1,line=4, "B~N(50,10), P~N(0.1,0.02)")
for(i in 1:21){
  MSDP[i]<-mean(SDP[,i])
}
for(i in 1:21){
  MSDP2[i]<-mean(SDP2[,i])
}
plot(r,MSDP,type="l",xlab="Correlation Coefficient r",ylab="SD(P)")
lines(r,MSDP2,type="l",lty=2,col=2)
legend(0,0.15,c("Before Transformation",
  "After Transformation"),lty=c(1,2),col=c(1,2),bty="n")
mtext(side=1,line=4, "B~N(50,10), P~N(0.1,0.02)")

```