

# Does more data mean higher efficiency? An experience from pre- and post-treatment study with missing data

Hongyue WANG<sup>1</sup>, Jing PENG<sup>1</sup>, Juila Z. ZHENG<sup>3</sup>, Bokai WANG<sup>1</sup>, J. X. TU<sup>4</sup>, Changyong FENG<sup>1,2,\*</sup>

**Summary:** In this paper we compare two moment-based methods which have been widely used to test the hypothesis of no treatment effect in pre- and post-treatment studies with data missing completely at random. Our theoretical derivation and simulation results show that the method based on all available data is not necessarily more efficient than the method that uses only complete data pairs. We propose an optimal linear combination of these two methods which turns to be more powerful in all cases.

**Key words:** paired t-test; likelihood ratio test; asymptotical relative efficiency

[Shanghai Arch Psychiatry. 2016; 28(4): 235-240. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.216058>]

## 1. Introduction

It is well known in data analysis that more data usually offer more information to make statistical inferences. For example, suppose we want to find the average body weight of 2-year-old boys in New York City. For this purpose, we can randomly select 100 boys from the targeted population, obtain their individual body weight, and calculate the average body weight and the standard deviation. If possible, we can also randomly select 10,000 boys and do the same calculations. Usually, the average weights in both cases are very similar. However, the standard deviation of the latter is only about 10% of the former one.

Student's t-test<sup>[1]</sup> is one of the most popular statistical tools used to compare the mean value with continuously distributed data. Let  $X$  be a random sample from a population of interest with mean value  $\mu$  and variance  $\sigma^2$ . The sample mean and sample variance are

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

which are unbiased estimators of  $\mu$  and  $\sigma^2$ , respectively<sup>[2]</sup>. A widely used method to test the hypothesis  $H_0: \mu = \mu_0$  is the test statistic defined by

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}.$$

If the data follows normal distribution, then under the null hypothesis  $H_0$ , the test statistic  $T$  in (1) has a  $t$ -distribution with  $n-1$  degrees of freedom<sup>[3]</sup>. If the data is not normally distributed, the exact distribution of  $T$  is usually not available. However, as long as the sample size  $n$  is large enough, we can use the standard normal distribution to approximate the distribution of  $T$ , which is the direct result of the central limit theorem in probability theory<sup>[4]</sup>. The test statistic in (1) is also called the one-sample  $t$ -test in statistics<sup>[3]</sup>.

Now consider the case of two independent samples. Suppose  $X_{i1}, X_{i2}, \dots, X_{in}$ ,  $i=1,2$  are data from two independent populations with means  $\mu_i$  and variances  $\sigma_i^2$ ,  $i=1, 2$ . Let  $\bar{X}_i$  and  $S_i^2$  be the sample means and

<sup>1</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY

<sup>2</sup>Department of Anesthesiology, University of Rochester, Rochester, NY

<sup>3</sup>Department of Microbiology and Immunology, McGill University, Montreal, QC

<sup>4</sup>State University of New York Upstate Medical University, Syracuse, NY

\*correspondence: Dr. Changyong Feng. Mailing address: Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Ave., Box 630, Rochester, NY, USA. Postcode: NY 14642. E-mail: [Changyong\\_feng@urmc.rochester.edu](mailto:Changyong_feng@urmc.rochester.edu)

sample variances in these two samples. The widely used test statistic for testing the hypothesis is  $H_0: \mu_1 = \mu_2$  is

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

If data are normally distributed in both samples and  $\sigma_1^2 = \sigma_2^2$ , under  $H_0$ ,  $T$  in (2) has a  $t$ -distribution with degrees of freedom  $n_1 + n_2 - 2$ . If  $\sigma_1^2 \neq \sigma_2^2$ , the distribution of  $T$  in (2) is not so straightforward. This is the well-known Behrens-Fisher problem<sup>[3]</sup> in statistics and is out of the scope of this paper. However, if both  $n_1$  and  $n_2$  are large enough, we can still use the standard normal distribution to approximate the distribution of  $T$ . The test statistic in (2) is also called the two-sample  $t$ -test in statistics<sup>[3]</sup>. In two-sample  $t$ -test, the groups usually have different sample sizes.

Consider another scenario. Suppose we have a set of randomly selected, matched pair observations  $(X_{i1}, X_{i2}), i=1, \dots, n$ , from a study population. This kind of data is very typical in pre- and post- treatment study. For example, in a hypertension study,  $X_{i1}$  and  $X_{i2}$  are the blood pressure of patient  $i$  before and after the treatment. This is different from the two independent samples considered above. For matched pair data,  $X_{i1}$  and  $X_{i2}$  are correlated as they are two measurements on the same individual. Suppose in the study population, the mean blood pressures before and after the treatment are  $\mu_1$  and  $\mu_2$ , respectively. The treatment effect can be measured by  $\mu_1 - \mu_2$ . Let  $Y_i = X_{i1} - X_{i2}$ , the difference of measurements before and after treatment. Let  $\bar{Y}$  and  $S_Y^2$  be the sample mean and sample variance of  $Y_i, i=1, \dots, n$ . The test statistic widely used in statistics to test the hypothesis  $H_0: \mu_1 = \mu_2$  is

$$T = \frac{\sqrt{n}\bar{Y}}{S_Y} = \frac{\sqrt{n}(\bar{X}_1 - \bar{X}_2)}{S_Y}$$

If  $(X_{i1}, X_{i2})$  has a bivariate distribution<sup>[2]</sup>, the test statistic in (3) has  $t$ -distribution with degrees of freedom  $n - 1$ , and is called the paired  $t$ -test in statistics<sup>[3]</sup>.

Note that the two-sample  $t$ -test and the paired  $t$ -test can be written in the same form

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\widehat{\text{Var}}(\bar{X}_1 - \bar{X}_2)}}$$

where  $\widehat{\text{Var}}(\bar{X}_1 - \bar{X}_2)$  is an  $\sqrt{n}$  consistent estimator<sup>[3]</sup> of the variance of  $\bar{X}_1 - \bar{X}_2$ . The formula of  $\widehat{\text{Var}}(\bar{X}_1 - \bar{X}_2)$  obtained from the authors upon request. From (3) we know that the paired  $t$ -test is exactly the one-sample  $t$ -test constructed from the pre- and post-treatment differences.

In this paper we focus on the matched pair data. In the construction of test statistic (3) we assume that the pre- and post-treatment data are available for each individual. The power of the test increases with the sample size. However, missing data is very typical for pre- and post-treatment studies. Usually the pre-

treatment measurement is available for each individual. However, the post-treatment data may be missing for some individuals. This puts some challenges on the data analysis. For example, to test the hypothesis of no treatment effect, we may construct the test statistics of the same structure as (4). However, for the estimation of the mean value of the pre-treatment measurements, should we use all individuals, or only the individuals with complete pairs? What is the relative efficiency of the test statistics based these two different estimations?

The paper is organized as follows. Section 2 introduces two widely used moment-based test statistics, and calculates their relative efficiency. In Section 3 we construct a test which is more powerful than the tests in Section 2, and is equivalent to the likelihood ratio test when the data is from bivariate normal distribution. In Section 4, we do some simulation studies to compare the powers of these tests. Our conclusion and further discussion are reported in Section 5.

## 2. Two moment-based tests and their relative efficiency

Suppose the full data is  $(X_{i1}, X_{i2}), i=1, \dots, n$ , where  $X_{i1}$  and  $X_{i2}$  are pre- and post-treatment measurements, respectively. The pre-treatment measurement is observed for every individual. However, for some individuals, the post-treatment measurements are not observed. For individual  $i$ , we define an indicator  $R_i$ , with  $R_i=1$  if  $X_{i2}$  is observed, and  $R_i=0$  otherwise. Hence the number of complete pairs (*i.e.* both pre- and post-treatment measurements are observed) is  $\sum_{i=1}^n R_i$ . Data can be missing in very complicated patterns in biomedical research, especially in longitudinal follow-up studies. See Rubin<sup>[5]</sup>, Little and Rubin<sup>[6]</sup> for more theoretical discussions about the missing patterns. In this manuscript, we consider a very simple missing pattern where the post-treatment is assumed to be missing completely at random (MCAR)<sup>[5,6]</sup>, which means that the probability that  $X_{i2}$  is missing does not depend on  $X_{i1}$ . This is a strong assumption. For example, suppose  $X_{i1}$  and  $X_{i2}$  are the blood pressures before and after the treatment. If the patient skips the post-treatment because he accidentally forgets the appointment, the MCAR assumption is satisfied. However, if the patient thinks his blood pressure at the first appointment is in the normal range and he doesn't want to waste time to do the second measurement, the MCAR assumption is invalid in this case as the missing depends on the first measurement.

We assume the mean and variance of  $X_{ij}$  are  $\mu_j$  and  $\sigma_j^2, j=1, 2$ . Since  $X_{i1}$  and  $X_{i2}$  are from the same individual, they are usually correlated (assuming their correlation coefficient is  $\rho$ ). Given the data, these parameters need to be estimated in order to make appropriate statistical inference. With MCAR data,  $\sqrt{n}$ -consistent estimators can be easily obtained for all these parameters.

For the pre- and post-treatment data, we are interested in the treatment effect which can be measured by  $\mu_1 - \mu_2$ . Statistically, the hypothesis for no treatment effect is the same as  $H_0: \mu_1 = \mu_2$ .

**2.1 Test statistic based on all available data**

From formula (4) we know that test statistic depends on the estimation of the pre- and post- treatment means and the (estimator of) the variance of the sample mean difference. In this section, the sample mean (denoted by  $\hat{\mu}_{1A}$ ) of the pre-treatment measurement is calculated based on all individuals, and the sample mean (denoted by  $\hat{\mu}_2$ ) of the post-treatment is calculated only based on the post-treatment measurements, i.e.

$$\hat{\mu}_{1A} = \frac{1}{n} \sum_{i=1}^n X_{i1}, \hat{\mu}_2 = \frac{\sum_{i=1}^n R_i X_{i2}}{\sum_{i=1}^n R_i}$$

The *t*-test based on all available data is

$$T_A = \frac{\sqrt{n}(\hat{\mu}_{1A} - \hat{\mu}_2)}{\sqrt{\widehat{\text{Var}}(\hat{\mu}_{1A} - \hat{\mu}_2)}}$$

The exact distribution of  $T_A$  is difficult to calculate. However, under MCAR, when sample size *n* is large enough, the normal distribution can be used to approximate the asymptotic distribution of  $T_A$ <sup>[3]</sup>.

**2.2 Test statistic based on complete pairs**

In this method, the sample mean (denoted by  $\hat{\mu}_{1C}$ ) of the pre-treatment measurement is only based on individuals with complete pairs, i.e.

$$\hat{\mu}_{1C} = \frac{\sum_{i=1}^n R_i X_{i1}}{\sum_{i=1}^n R_i}$$

The *t*-test is exactly the paired *t*-test

$$T_C = \frac{\sqrt{n}(\hat{\mu}_{1C} - \hat{\mu}_2)}{\sqrt{\widehat{\text{Var}}(\hat{\mu}_{1C} - \hat{\mu}_2)}}$$

Similarly, under MCAR, the asymptotic distribution of  $T_C$  can be approximated by normal distribution for large sample size.

**2.3 Asymptotic relative efficiency**

The relative efficiency of two tests is used to characterize their powers<sup>[3]</sup>. Let  $r_\sigma = \sigma_1 / \sigma_2$ , the ratio of the standard deviations of the pre- and post-treatment measurements. Let  $\pi$  denote the probability that the post-treatment measurement is observed. It can be proved that the asymptotic relative efficiency of  $T_C$  with respect to  $T_A$  is

$$ARE_{CA} = \pi + \frac{1 - \pi}{(1 - \rho^2) + (\rho - r_\sigma)^2} \tag{5}$$

$$= 1 + \frac{(1 - \pi)r_\sigma(2\rho - r_\sigma)}{(1 - \rho^2) + (\rho - r_\sigma)^2} \tag{6}$$

Here  $ARE_{CA} > (or <) 1$  means that  $T_C$  is more (or less) powerful than  $T_A$  to detect the pre- and post- treatment difference if it exists. From formula (5) we can see that the relative efficiency depends on the proportion of missing ( $1 - \pi$ ), the ratio of the variances of the pre- and post-treatment measurements, and their correlation. More specifically, we have the following conclusions about  $ARE_{CA}$ :

- (i) Formula (5) shows that  $ARE_{CA}$  is always greater than  $\pi$ . This is very intuitive as  $\pi$  is the proportion of patients without missing data.
- (ii) If  $r_\sigma \geq 2$ ,  $T_A$  is more powerful than  $T_C$ .
- (iii) If  $r_\sigma / 2 \leq \rho < 1$ ,  $T_C$  is more powerful than  $T_A$ .
- (iv) If  $-1 \leq \rho < r_\sigma / 2 < 1$ ,  $T_A$  is more powerful than  $T_C$ .
- (v) If  $r_\sigma 1$ , i.e.  $\sigma_1 = \sigma_2$ , then  $\rho > 0$ . This means that for highly (positively) correlated data,  $T_C$  is more powerful than  $T_A$ .

It is interesting to see that  $T_A$  is not always more powerful than  $T_C$ , as one would have expected since the former test is based on more data than the latter one. When  $\sigma_1 < 2\sigma_2$ ,  $T_C$  is actually more efficient than  $T_A$  if  $r_\sigma / 2 \leq \rho \leq 1$ . In addition, in the special case of  $\sigma_1 = \sigma_2$ ,  $T_C$  can be much more efficient than  $T_A$  if the pre- and post-treatment measurements are highly correlated.

**3. An optimal combination of moment-based tests**

Section 2 shows that although  $T_A$  and  $T_C$  are the same when data is not missing; none of them is uniformly more powerful than the other when data is missing completely at random. A very intuitive idea to find an intermediate point between those two tests which may be at least as powerful as both of them. More precisely, consider the following set

$$F = \{T(\lambda) = \lambda T_A + (1 - \lambda)T_C : 0 \leq \lambda \leq 1\} \tag{7}$$

Each element in  $F$  is a valid test, and  $T_A$  and  $T_C$  are two special elements in this family.

**Theorem 1.** Among all tests defined in (7),  $T(\lambda_o)$  is the most powerful one, where

$$\lambda_o = \frac{a - b}{(1 + a)(1 - b)},$$

and

$$a = \sqrt{\frac{r_{\sigma}^2 - 2\rho r_{\sigma} + 1}{\pi r_{\sigma}^2 - 2\rho\pi r_{\sigma} + 1}}$$

$$b = \frac{\pi r_{\sigma}^2 - (1 + \pi)\rho r_{\sigma} + 1}{\sqrt{(r_{\sigma}^2 - 2\rho r_{\sigma} + 1)(\pi r_{\sigma}^2 - 2\rho\pi r_{\sigma} + 1)}}$$

The proof of this theorem is out of the scope of this paper, but it is available from the authors upon request.

**Remark:** It is well known that if the data is from bivariate normal distribution, the likelihood ratio test (LRT) is the most efficient test<sup>[3]</sup>. We can prove that  $T(\lambda_0)$  in Theorem 1 is equivalent to the likelihood ratio test for bivariate normal data. It only depends on the first two moments of the data, is easy to use, and is more powerful than currently widely used two tests TA and TC. Same idea of combination has been used in

other area of statistics. For example, Oakes and Feng<sup>[7]</sup> constructed of an optimal log linear combination of the stratified and unstratified log-rank tests<sup>[8]</sup>.

#### 4. Simulation results

In this section we compare the empirical power of  $T_A$ ,  $T_C$  and  $T(\lambda_0)$  for different sample sizes and different parameters in the distribution of the data. The significance level was set at 5% for all cases. About 30% of the post-treatment data is missing. For each test statistic  $T$ , we first standardize it to make its (asymptotic) variance equal 1. The empirical power is obtained from 10, 000 Monte Carlo replications. The empirical power is the proportion of times that  $|T| > 1.96$ .

##### Case 1. Bivariate normal data

In this case, the matched pair  $(X_{i1}, X_{i2})$  are generated from bivariate normal distribution<sup>[21]</sup>. We report the powers of  $T_A$ ,  $T_C$ ,  $T(\lambda_0)$  and LRT. The result is in Table 1.

**Table 1: Comparison of powers of test statistics (bivariate normal data)**

n	Parameters					Powers of test statistics			
	$\mu_1$	$\mu_2$	$\sigma_1$	$\sigma_2$	$\rho$	$T_A$	$T_C$	$T(\lambda_0)$	LRT
50	0.0	0.5	3.0	1.0	0.6	0.27	0.21	0.27	0.27
	0.0	0.5	1.0	2.0	0.3	0.32	0.42	0.42	0.42
	0.0	0.5	1.0	2.0	0.6	0.39	0.44	0.44	0.44
	0.0	0.5	2.0	1.0	0.6	0.52	0.44	0.54	0.54
	0.0	0.25	1.0	1.0	0.9	0.60	0.91	0.91	0.91
	0.0	0.25	0.6	1.0	0.3	0.31	0.31	0.32	0.32
	0.0	0.25	0.6	1.0	0.6	0.40	0.45	0.45	0.45
	0.0	0.25	1.0	1.0	0.8	0.49	0.65	0.66	0.66
100	0.0	0.5	3.0	1.0	0.6	0.48	0.38	0.49	0.49
	0.0	0.5	1.0	2.0	0.3	0.57	0.57	0.58	0.58
	0.0	0.5	1.0	2.0	0.6	0.67	0.74	0.74	0.74
	0.0	0.5	2.0	1.0	0.6	0.82	0.74	0.84	0.84
	0.0	0.25	1.0	1.0	0.9	0.88	1.00	1.00	1.00
	0.0	0.25	0.6	1.0	0.3	0.55	0.55	0.56	0.56
	0.0	0.25	0.6	1.0	0.6	0.67	0.74	0.74	0.74
	0.0	0.25	1.0	1.0	0.8	0.78	0.91	0.92	0.92
200	0.0	0.5	3.0	1.0	0.6	0.77	0.65	0.78	0.78
	0.0	0.5	1.0	2.0	0.3	0.85	0.85	0.87	0.87
	0.0	0.5	1.0	2.0	0.6	0.93	0.95	0.95	0.95
	0.0	0.5	2.0	1.0	0.6	0.98	0.95	0.99	0.99
	0.0	0.25	1.0	1.0	0.9	0.99	1.00	1.00	1.00
	0.0	0.25	0.6	1.0	0.3	0.84	0.84	0.85	0.85
	0.0	0.25	0.6	1.0	0.6	0.93	0.96	0.96	0.96
	0.0	0.25	1.0	1.0	0.8	0.97	1.00	1.00	1.00

- (1) As expected, given the parameters in the distribution of the data, the power of each test increases with the sample size.
- (2) For  $T_A$  and  $T_C$ , none of them is always more powerful than the other. For example, given sample size  $n=200$ , when  $\mu_1=0$ ,  $\mu_2=0.5$ ,  $\sigma_1=3.0$ ,  $\sigma_2=1.0$ , and  $\rho=0.6$ , the powers of  $T_A$  and  $T_C$  are 0.77 and 0.65, respectively. However, when  $\mu_1=0$ ,  $\mu_2=0.5$ ,  $\sigma_1=1.0$ ,  $\sigma_2=2.0$ , and  $\rho=0.6$ , their powers are 0.93 and 0.95, respectively.
- (3) In any scenario,  $T(\lambda_o)$  is more powerful than  $T_A$  and  $T_C$  even when sample size is relatively small (e.g.  $n=50$ ), and it always has the same power as the likelihood ratio test.

*Case 2. Mixed normal-exponential data*

The data is generated in the following form:

$$X_{i1} = X_{i11} + X_{i12}$$

$$X_{i2} = X_{i21} + X_{i22}$$

where  $X_{i11}$ ,  $X_{i21}$ , and  $X_{i12}$  are independent random variables;  $X_{i11}$  and  $X_{i21}$  have normal distribution, and  $X_{i12}$  has exponential distribution. In this case, the data does not have bivariate normal distribution. However, as long as the sample size is large enough, we can still use the t-test to compare the pre- and post-treatment mean values. Table 2 reports the empirical powers of  $T_A$ ,  $T_C$ , and  $T(\lambda_o)$ . It shows that neither  $T_A$  nor  $T_C$  is more powerful than the other in all situations. However,  $T(\lambda_o)$  is always more powerful than both of them.

**Table 2. Comparison of powers of test statistics (mixed normal-exponential data)**

n	Parameters					Powers of test statistics		
	$\mu_1$	$\mu_2$	$\sigma_1$	$\sigma_2$	$\rho$	$T_A$	$T_C$	$T(\lambda_o)$
50	0.0	1.5	6.25	4.0	0.6	0.47	0.42	0.49
	0.0	1.5	4.0	5.0	0.6	0.52	0.58	0.59
	0.0	1.5	4.0	5.0	0.3	0.39	0.38	0.40
	0.0	1.5	5.0	4.0	0.6	0.59	0.57	0.63
	0.0	1.0	4.0	4.0	0.9	0.61	0.91	0.91
	0.0	1.0	2.5	4.0	0.6	0.40	0.46	0.46
	0.0	1.0	2.5	4.0	0.3	0.31	0.31	0.32
	0.0	1.0	4.0	4.0	0.8	0.50	0.64	0.66
100	0.0	1.5	6.25	4.0	0.6	0.76	0.71	0.79
	0.0	1.5	4.0	5.0	0.6	0.81	0.86	0.86
	0.0	1.5	4.0	5.0	0.3	0.67	0.65	0.68
	0.0	1.5	5.0	4.0	0.6	0.87	0.86	0.90
	0.0	1.0	4.0	4.0	0.9	0.88	1.00	1.00
	0.0	1.0	2.5	4.0	0.6	0.68	0.74	0.74
	0.0	1.0	2.5	4.0	0.3	0.55	0.55	0.56
	0.0	1.0	4.0	4.0	0.8	0.79	0.91	0.92
200	0.0	1.5	6.25	4.0	0.6	0.97	0.95	0.97
	0.0	1.5	4.0	5.0	0.6	0.98	0.99	0.99
	0.0	1.5	4.0	5.0	0.3	0.92	0.91	0.93
	0.0	1.5	5.0	4.0	0.6	0.99	0.99	1.00
	0.0	1.0	4.0	4.0	0.9	0.99	1.00	1.00
	0.0	1.0	2.5	4.0	0.6	0.93	0.96	0.96
	0.0	1.0	2.5	4.0	0.3	0.84	0.83	0.84
	0.0	1.0	4.0	4.0	0.8	0.97	1.00	1.00

## 5. Conclusion

In the pre- and post-treatment studies, if the data is missing completely at random, we can construct test statistics either using all available data or using the complete pairs. These two methods only use the first two moments of the data and are very easy to implement. However, none of these two methods is uniformly better than the other in all cases. The relative efficiency of these two methods depends on the proportion of missing, the ratio of the variances, and the correlation of two measurements on the same individuals. In this paper, we propose a data-based method which is more powerful than those two methods. In fact, it is the most efficient test when data has bivariate normal distribution.

Missing data is a typical problem in pre- and post-treatment studies, and the missing pattern may be very complicated. MCAR is an over-simplified assumption. Another more realistic and still mathematically

tractable missing pattern is missing at random (MAR). Generalizing our method to MAR data is in progress.

## Acknowledgements

This study was supported by a pilot grant (PI: Feng) from the Clinical and Translational Sciences Institute at the University of Rochester Medical Center.

## Conflict of interest statement

The authors report no conflict of interest related to this manuscript.

## Author's contribution

Hongyue Wang, Bokai Wang, and Changyong Feng: theoretical derivation and revision. Julia Zheng, Jing Peng, and J. X. Tu: Simulation manuscript drafting

## 更多的数据意味着更高的效率吗？治疗前后缺失数据的研究经验

H WANG, J PENG, JZ ZHENG, B WANG, JX TU, C FENG

概述：在本文中，我们比较两个以时间为基础的方法，这两个方法已被广泛用于治疗前后随机缺失数据的研究中来测试没有治疗效果的假设。我们的理论推导和模拟结果表明，基于所有可用的数据的方法并不比使

用完整配对数据更有效。我们提出了一个合并这两个方法最佳的线性组合使其在所有案例中更有效。

关键词：配对 t 检验，似然比检验，渐近相对效率

## References

1. Student. The probable error of a mean. *Biometrika*. 1908; 6(1): 1–25
2. Lehmann EL. *Theory of Point Estimation*. New York: Springer-Verlag; 1983
3. Lehmann EL. *Testing Statistical Hypotheses* (2nd ed). New York: Springer-Verlag; 1982
4. Durrett R. *Probability: Theory and Examples* (4th ed). New York: Cambridge University Press; 2010
5. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63(3): 581–592. doi: <http://dx.doi.org/10.2307/2335739>
6. Little RJA, Rubin DR. *Statistical Analysis with Missing Data* (2nd ed). New York: Wiley; 2002
7. Oakes D, Feng C. Combining stratified and unstratified log-rank tests in paired survival data. *Stat Med*. 2010; 29(16):1735–1745. doi: <http://dx.doi.org/10.1002/sim.3921>
8. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. New York: Springer; 1993



Dr. Hongyue Wang obtained her BS in Scientific English from the University of Science and Technology of China (USTC) in 1995, and PhD in Statistics from the University of Rochester in 2007. She is a Research Associate Professor in the Department of Biostatistics and Computational Biology at the University of Rochester Medical Center. Her research interests include longitudinal data analysis, missing data, survival data analysis, and design and analysis of clinical trials. She has extensive and successful collaboration with investigators from various areas, including Infectious Disease, Nephrology, Neonatology, Cardiology, Neurodevelopmental and Behavioral Science, Radiation Oncology, Pediatric Surgery, and Dentistry. She has published more than 70 statistical methodology and collaborative research papers in peer-reviewed journals.