

Chapter 487

Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design

Introduction

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are sampled, rather than individual subjects. This sample size and power procedure is used for the case where the subject responses are counts (event rates). To reduce the variation (and thus increase power), clusters are matched, with one cluster of each pair assigned to the control group, and the other assigned the treatment group. This procedure gives the number of pairs needed for the desired power requirement.

The formulas used here are based on Hayes and Bennett (1999) and Hayes and Moulton (2009). The methods are discussed in Donner and Klar (2000) and briefly in Campbell and Walters (2014).

Technical Details

Definition of Terms

The following table presents the various terms that are used.

Group	1 (Control)	2 (Treatment)
Sample size:	K	K
Individual event rates:	λ_1	λ_2

Coefficient of Variation: CV_M (within-pair coefficient of variation between clusters in the absence of intervention)

Number of cluster pairs: K ($2K$ is the total number of clusters)

Person-Years per cluster: M

Hypotheses

The null and alternative hypotheses are

$$H_0: \lambda_{Diff} = 0 \quad \text{vs.} \quad H_1: \lambda_{Diff} \neq 0$$

Corresponding one-sided hypotheses may also be used. These hypotheses may be tested based on an appropriate paired difference test.

Sample Size and Power Calculations

Sample Size Calculation

The sample size calculation for the number of cluster pairs, as given in Hayes and Bennett (1999) and Hayes and Moulton (2009), is

$$K = 2 + (z_{\alpha/2} + z_{\beta})^2 \frac{(\lambda_1 + \lambda_2)/M + CV_M^2(\lambda_1^2 + \lambda_2^2)}{(\lambda_1 - \lambda_2)^2}$$

Estimating CV_M

Hayes and Bennett (1999) suggest that CV_M may be estimated based on the following, if prior data is available:

Let λ_{ij} represent the true event rate in the j^{th} cluster ($j = 1, 2$) of the i^{th} pair ($i = 1, \dots, M$) and r_{ij} represent the corresponding observed rate. The empirical variance of the i^{th} pair is given by

$$s_i^2 = (r_{i2} - r_{i1})^2 / 2$$

Define

$$s_m^2 = \sum s_i^2 / M$$

as the average of the within-pair variances.

CV_M may be estimated from

$$CV_M^2 = \frac{s_m^2 - Av(r_i/y_{ij})}{Av(r_i^2)}$$

where y_{ij} is the person-year of follow-up in the i^{th} pair of the j^{th} cluster, and $Av()$ indicates the mean over all M clusters.

According to Hayes and Bennett (1999), "If only unmatched data are available, a conservative approach is to use [the coefficient of variation (SD/Mean) between clusters within each group] as an upper limit for CV_M ."

If no data is available, a series of plausible values, usually between 0 and 0.5, should be considered.

Power Calculation

The corresponding power calculation to the sample size calculation above is

$$Power = 1 - \Phi \left(\sqrt{\frac{(K-2)(\lambda_1 - \lambda_2)^2}{(\lambda_1 + \lambda_2)/M + CV_M^2(\lambda_1^2 + \lambda_2^2)}} - z_{\alpha/2} \right)$$

where $z_x = \Phi(x)$ is the standard normal distribution function.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power*, *K*, *M*, and λ^2 .

Under most situations, you will select either *Power* to calculate power or *K* to calculate the number of cluster pairs. Occasionally, you may want to fix the number of clusters and find the necessary cluster size.

The value selected here always appears as the vertical axis on the charts (when applicable).

When searching for *M*, some scenarios with small *K*'s are not feasible.

Test

Alternative Hypothesis

Specify whether the test is one-sided or two-sided. The one-sided option specifies a one-tailed test.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (beta = 0.20) was used for power. Now, 0.90 (beta = 0.10) is commonly used.

A single value may be entered or a range of values, such as *0.8 to 0.95 by 0.05*, may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Usually, the value of 0.05 is used for two-sided tests and 0.025 is used for one-sided tests.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design

Sample Size – Number of Clusters & Cluster Size

K (Number of Cluster Pairs)

This is the number of cluster pairs. Each cluster pair consists of two clusters that are matched or paired based on one or more criteria. The total sample size is equal to the number of clusters times the person-years per cluster. This value must be a positive number. You can enter a single value such as 20 or a series of values such as 10 20 30 or 10 to 30 by 5.

M (Person-Years per Cluster)

This is the average number of person-years per cluster. This value must be a positive number that is at least 1. It can be a decimal (fractional) number such as 2.7. The name person-years is used for this parameter. However, we could have used person-time or person-week or person-day. It refers to a count of the number of events over a fixed time period. The scale of λ_1 and λ_2 should be in consistent with the Person-Years scale. λ_1 and λ_2 represent rates over a specific duration. You can enter a single value such as 50 or a series of values such as 40 50 60 or 40 to 60 by 5.

Effect Size

λ_1 (Event Rate of Group 1)

Enter a value (or range of values) for the mean event rate per unit time in group 1 (control group). The value must be greater than zero.

Example of Estimating λ_1

If 200 patients were exposed for 1 year and 40 experienced the event of interest, then the mean event rate would be

$$\lambda_1 = 40/(200*1) = 0.2 \text{ per patient-year}$$

If 200 patients were exposed for 2 years and 40 experienced the event of interest, then the mean event rate would be

$$\lambda_1 = 40/(200*2) = 0.1 \text{ per patient-year}$$

Event Rate Difference

λ_1 is used with λ_2 to calculate the event rate difference as

$$Diff = \lambda_2 - \lambda_1$$

such that

$$\lambda_1 = \lambda_2 - Diff$$

The range of acceptable values is $\lambda_1 > 0$. You can enter a single value such as 1 or a series of values such as 1 to 2 by 0.5.

Enter λ_2 , Diff, or Ratio for Group 2

This option lets you indicate how λ_2 is entered. The options are

- **λ_2 (Event Rate of Group 2)**
Enter the value of λ_2 directly.
- **Diff (Difference Between Event Rates)**
Enter values for the difference between the event rates ($Diff = \lambda_2 - \lambda_1$). The value of λ_2 is equal to $\lambda_1 + Diff$.

Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design

- **RR (Ratio of Event Rates)**

Enter values for the ratio of the event rates ($RR = \lambda_2 / \lambda_1$). The value of λ_2 is equal to $\lambda_1 \times$ Ratio. Note that the hypothesis still concerns the difference. This is just a convenient way of specifying a value.

λ_2 (Event Rate of Group 2)

This option is displayed only if Enter λ_2 , Diff, or Ratio for Group 2 = " λ_2 (Event Rate of Group 2)."

Enter a value (or range of values) for the mean event rate per time unit in group 2 (treatment group). The value must be greater than zero and different from λ_1 .

Example of Estimating λ_2

If 200 patients were exposed for 1 year and 40 experienced the event of interest, then the mean event rate would be

$$\lambda_2 = 40/(200*1) = 0.2 \text{ per patient-year}$$

If 200 patients were exposed for 2 years and 40 experienced the event of interest, then the mean event rate would be

$$\lambda_2 = 40/(200*2) = 0.1 \text{ per patient-year}$$

Event Rate Difference

λ_2 is used with λ_1 to calculate the event rate difference as

$$Diff = \lambda_2 - \lambda_1$$

such that

$$\lambda_2 = \lambda_1 + Diff$$

The range of acceptable values is $\lambda_2 > 0$. You can enter a single value such as 1 or a series of values such as 1 to 2 by 0.5.

Diff (Difference Between Event Rates)

This option is displayed only if Enter λ_2 , Diff, or Ratio for Group 2 = " $Diff$ (Difference Between Event Rates)."

Enter values for the difference between the event rates ($Diff = \lambda_2 - \lambda_1$). The value of λ_2 is equal to $\lambda_1 + Diff$. The range is any positive or negative value other than 0 ($Diff \neq 0$). You can enter a single value such as 0.5 or a series of values such as 0.5 0.6 0.7 or 1 to 6 by 1.

RR (Ratio of Event Rates)

This option is displayed only if Enter λ_2 , Diff, or Ratio for Group 2 = " RR (Ratio of Event Rates)."

Enter values for the ratio of the event rates ($RR = \lambda_2 / \lambda_1$) at which the power is to be calculated. The value of λ_2 is calculated using $\lambda_1 \times RR$. The range is $RR > 0$ and $RR \neq 1$. Typically, RR is between 0.5 and 2.0. You can enter a single value such as 1.15 or a series of values such as 1.05 1.10 1.15 1.20 or 1.05 to 1.20 by 0.05.

CVM (Within-Pair Coefficient of Variation)

Enter a value or values for the within-pair coefficient of variation (SD/Mean) between clusters in the absence of intervention. See the documentation or references for suggestions for estimating CVM. If prior information is not available, Hayes and Bennett (1999) suggest that CVM is usually less than 0.25 and seldom greater than 0.50. You can enter a single value such as 0.15 or a series of values such as 0.05 0.10 0.15 0.20 or 0.05 to 0.30 by 0.05.

Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design

Example 1 – Calculating Sample Size

One difficulty in calculating sample size for a matched-pair difference of two event rates in a cluster-randomized design is obtaining a value for CV_M , the within-pair coefficient of variation between clusters. This example shows how to enter a range of values to determine the effect of CV_M on sample size. Suppose that a cluster randomized study is to be conducted in which the control event rate is assumed to be $\lambda_1 = 0.80$, and the treatment event rate is $\lambda_2 = 0.58$. Each cluster is assumed to have about 80 individuals. Each will be followed for one year. A range of values between 0.05 and 0.5 for CV_M will be examined. The desired power and alpha are 0.9 and 0.05, respectively. The test will be a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	K (Number of Cluster Pairs)
Alternative Hypothesis	Two-Sided
Power	0.90
Alpha	0.05
M (Person-Years per Cluster)	80
λ_1 (Event Rate of Group 1)	0.80
Enter λ_2 , Diff, or Ratio for Group 2	λ_2 (Event Rate of Group 2)
λ_2 (Event Rate of Group 2)	0.58
CVM (Within-Pair Coefficient of Variation)	0.05 to 0.50 by 0.05

Annotated Output

Click the *Calculate* button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for a Two-Sided Test of Event-Rate Matched-Pair Difference in a Cluster-Randomized Design

Group 1 = Control. Group 2 = Treatment.

	Number of Cluster Pairs	Total Number of Clusters	Person- Years per Cluster	Total Person- Years N = 2KM	Event Rate Gr 1 λ_1	Event Rate Gr 2 λ_2	Event Rate Rate Diff. $\lambda_2 - \lambda_1$	Event Rate Ratio λ_2 / λ_1	Coeffi- cient of Variation CVM	Two- Sided Alpha
Power	K	2K	M							
0.9389	7	14	80	1120	0.8000	0.5800	-0.2200	0.73	0.050	0.050
0.9064	8	16	80	1280	0.8000	0.5800	-0.2200	0.73	0.100	0.050
0.9151	11	22	80	1760	0.8000	0.5800	-0.2200	0.73	0.150	0.050
0.9167	15	30	80	2400	0.8000	0.5800	-0.2200	0.73	0.200	0.050
0.9001	19	38	80	3040	0.8000	0.5800	-0.2200	0.73	0.250	0.050
0.9022	25	50	80	4000	0.8000	0.5800	-0.2200	0.73	0.300	0.050
0.9027	32	64	80	5120	0.8000	0.5800	-0.2200	0.73	0.350	0.050
0.9025	40	80	80	6400	0.8000	0.5800	-0.2200	0.73	0.400	0.050
0.9020	49	98	80	7840	0.8000	0.5800	-0.2200	0.73	0.450	0.050
0.9013	59	118	80	9440	0.8000	0.5800	-0.2200	0.73	0.500	0.050

Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design

References

- Hayes, R.J. and Bennett, S. 1999. 'Simple sample size calculation for cluster-randomized trials'. International Journal of Epidemiology. Vol 28, pages 319-326.
- Hayes, R.J. and Moulton, L.H. 2009. Cluster Randomised Trials. CRC Press. New York.
- Campbell, M.J. and Walters, S.J. 2014. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Wiley. New York.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

K represents the number of cluster pairs needed in the design.

2K is the total number of clusters in the design, 2 clusters per pair.

M is the average number of person-years per cluster in all clusters.

N is the total number of person-years in the design. $N = 2KM$.

λ_1 is the event (or incidence) rate of the control group. This is the baseline rate.

λ_2 is the event (or incidence) rate of the treatment group.

$\lambda_2 - \lambda_1$ is the difference between the treatment event rate and the control event rate.

λ_2 / λ_1 is the ratio of the treatment event rate to the control event rate.

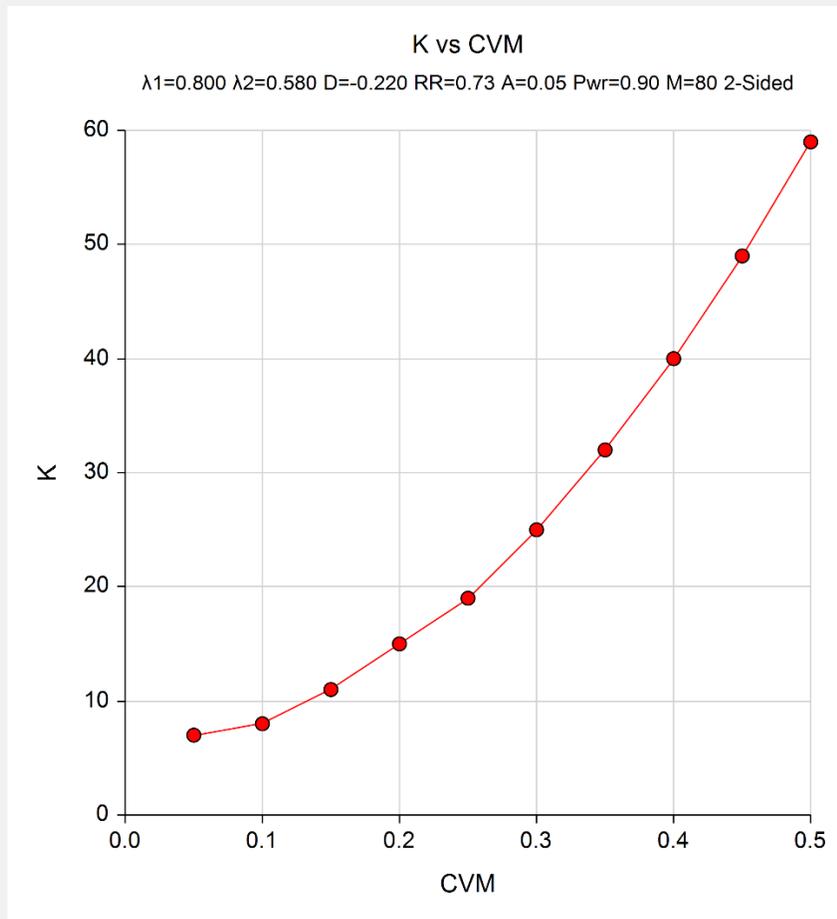
CVM is the within-pair coefficient of variation between clusters in the absence of intervention. See the documentation or references for suggestions for estimating CVM.

Alpha is the probability of rejecting a true null hypothesis, that is, rejecting H_0 when the event rates are actually equal.

Summary Statements

A sample of 7 cluster pairs (14 clusters) with 80 person-years per cluster achieves 94% power to detect a difference of -0.2200 between the treatment event rate 0.5800 and the control event rate 0.8000. The within-pair coefficient of variation between clusters is 0.050. A two-sided paired test of the event-rate difference is assumed, with a significance level of 0.050.

Chart Section



This report shows the needed number of cluster pairs for each of the coefficient of variation values.

Example 2 – Validation using Direct Calculation

Hayes and Bennett (1999) and Hayes and Moulton (2009) give formulas for sample size for tests of the matched-pair difference in a cluster-randomized design for comparing event rates, proportions, and means. However, the only example corresponds to the proportions case. Thus, we give an example where the number of cluster pairs needed is determined by direct calculation.

We will consider the (two-sided) case where $\lambda_1 = 0.6$; $\lambda_2 = 0.4$; $M = 200$; $\alpha = 0.05$; $CVM = 0.25$; $Power = 0.80$.

$$\begin{aligned}
 K &= 2 + \left(\frac{z_\alpha}{2} + z_\beta \right)^2 \frac{(\lambda_1 + \lambda_2)/M + CV_M^2(\lambda_1^2 + \lambda_2^2)}{(\lambda_1 - \lambda_2)^2} \\
 &= 2 + (1.9600 + 0.8416)^2 \frac{(0.6 + 0.4)/200 + 0.25^2(0.6^2 + 0.4^2)}{(0.4 - 0.6)^2} \\
 &= 2 + 7.84888 \frac{0.005 + 0.0625(0.52)}{0.04} = 9.358
 \end{aligned}$$

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Matched-Pair Difference of Two Event Rates in a Cluster-Randomized Design** procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	K (Number of Cluster Pairs)
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
M (Person-Years per Cluster)	200
λ_1 (Event Rate of Group 1)	0.60
Enter λ_2 , Diff, or Ratio for Group 2	λ_2 (Event Rate of Group 2)
λ_2 (Event Rate of Group 2)	0.40
CVM (Within-Pair Coefficient of Variation)	0.25

Output

Click the *Calculate* button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for a Two-Sided Test of Event-Rate Matched-Pair Difference in a Cluster-Randomized Design

Group 1 = Control. Group 2 = Treatment.

	Number of Cluster Pairs	Total Number of Clusters	Person-Years per Cluster	Total Person-Years	Event Rate Gr 1	Event Rate Gr 2	Event Rate Diff. $\lambda_2 - \lambda_1$	Event Rate Ratio λ_2 / λ_1	Coefficient of Variation CVM	Two-Sided Alpha
Power	K	2K	M	N = 2KM	λ_1	λ_2	$\lambda_2 - \lambda_1$	λ_2 / λ_1	CVM	Alpha
0.8318	10	20	200	4000	0.6000	0.4000	-0.2000	0.67	0.250	0.050

PASS calculates the number of cluster pairs needed to be 10, which matches the rounded-up calculation value.