Biostatistics – Spring 2026
Lecture 05: Statistical Comparison of Rates Between Groups

Dr. Zaid T. Al-Khaledi
Department of Statistics and Informatics
University of Mosul

March 11, 2026

Introduction

In previous lectures, we studied contingency tables and the main tests used for categorical data: the Chi-square test, Fisher's Exact Test, and Yates continuity correction.

In many biomedical and public health applications, the practical question is not only whether two variables are associated. Instead, we ask:

Are the disease rates, death rates, or event rates the same in two groups?

For example:

- Is the mortality rate the same in two time periods?
- Is the infection rate the same in males and females?
- Is the recovery rate the same in a treated group and a control group?

This lecture explains how rate comparison is written statistically, how the data are arranged in a 2×2 table, which test is appropriate, and how the result should be interpreted in biomedical language.

By the end of this lecture, you should be able to:

- Define and compute a simple event rate in each group.
- Formulate hypotheses for comparing two rates.
- Construct the corresponding 2×2 table.
- Decide whether to use Chi-square, Yates correction, or Fisher's Exact Test.
- Interpret the result statistically and medically.

1. What Do We Mean by Comparing Rates?

Suppose a population is divided into two groups:

$$n_1 \quad \text{and} \quad n_2$$

and suppose the event of interest occurred in

$$a_1 \quad \text{and} \quad a_2$$

individuals in the two groups, respectively.

Then the two observed rates are:

$$r_1 = \frac{a_1}{n_1}, \qquad r_2 = \frac{a_2}{n_2}.$$

These are often called event rates, disease rates, death rates, or proportions, depending on the context.

Example

If 12 out of 80 patients in Ward A developed infection, and 5 out of 70 patients in Ward B developed infection, then

$$r_1 = \frac{12}{80} = 0.15$$
$$r_2 = \frac{5}{70} \approx 0.0714.$$

Interpretation:

The observed infection rate is 15% in Ward A and about 7.14% in Ward B.

2. Hypotheses for Comparing Two Rates

Usually, we want to test whether the two population rates are equal.

2.1 Null hypothesis

$$H_0 : \frac{a_1}{n_1} = \frac{a_2}{n_2}$$

or equivalently,

$$H_0 : \frac{a_1}{n_1} - \frac{a_2}{n_2} = 0.$$

2.2 Alternative hypothesis

For a two-sided test:

$$H_A : \frac{a_1}{n_1} \neq \frac{a_2}{n_2}.$$

For a one-sided test, if we want to check whether group 1 has a larger rate:

$$H_A : \frac{a_1}{n_1} > \frac{a_2}{n_2}.$$

If we want to check whether group 1 has a smaller rate:

$$H_A : \frac{a_1}{n_1} < \frac{a_2}{n_2}.$$

Important note

In most medical and epidemiological examples in this course, we use the **two-sided test** unless the study gives a strong reason for a one-sided alternative.

3. Arranging the Data in a 2×2 Table

The comparison of two rates can be written as a 2×2 table:

	Event	No Event	Total
Group 1	a_1	b_1	n_1
Group 2	a_2	b_2	n_2
Total	A	B	N

where

$$b_1 = n_1 - a_1, \qquad b_2 = n_2 - a_2,$$

$$A = a_1 + a_2, \qquad B = b_1 + b_2, \qquad N = n_1 + n_2.$$

This table is exactly the same structure used in Chi-square and Fisher tests.

Why is this useful?

Because once the data are written in this table, the problem of comparing rates becomes a problem of testing association in a 2×2 table.

4. Which Test Should We Use?

Since you already studied the main tests, here is the practical decision rule.

4.1 Chi-square test

Use the ordinary Chi-square test when:

- sample size is reasonably large,
- expected counts are not too small,
- especially when all expected counts are at least 5.

4.2 Yates continuity correction

Use Yates correction when:

- the table is 2×2,
- sample size is not very large,
- but it is not extremely small.

Yates correction makes the test more conservative.

4.3 Fisher's Exact Test

Use Fisher's Exact Test when:

- sample size is small,
- or one or more expected counts are below 5,
- or the table contains very small observed values such as 0, 1, or 2.

Simple summary

Large sample → Chi-square.
Borderline small 2×2 table → Yates correction.
Very small sample or tiny expected count → Fisher's Exact Test.

5. Comparing Two Rates Using Chi-square

When the sample is large enough, the difference between two rates can be tested through the 2×2 table.

The rates are

$$r_1 = \frac{a_1}{n_1}, \qquad r_2 = \frac{a_2}{n_2}$$

and the pooled proportion is

$$p = \frac{a_1 + a_2}{n_1 + n_2} = \frac{A}{N}, \qquad q = 1 - p = \frac{B}{N}.$$

One common large-sample form is

$$\chi^2 = \frac{(a_1 n_2 - a_2 n_1)^2 N}{A B n_1 n_2}.$$

For a 2×2 table, the degrees of freedom are

$$df = 1.$$

Decision rule at significance level $\alpha = 0.05$:

$$\text{Reject } H_0 \text{ if } \chi^2_{\text{cal}} > \chi^2_{0.05,1} = 3.841.$$

6. Example 1: Comparison of Mortality Rates in Two Periods

Suppose the following data describe deaths among infected patients in two time periods.

Period	Number of infected (n_i)	Number of deaths (a_i)
1994–2000	40	10
2001–2007	61	5

Step 1: Compute the rates

$$r_1 = \frac{10}{40} = 0.25$$

$$r_2 = \frac{5}{61} \approx 0.082.$$

If we express them per 1000 infected persons:

$$0.25 \times 1000 = 250$$

$$0.082 \times 1000 \approx 82.$$

Interpretation: The mortality rate decreased from 250 per 1000 infected persons in the first period to about 82 per 1000 infected persons in the second period.

Step 2: State the hypotheses

$$H_0 : \frac{10}{40} = \frac{5}{61}$$

versus

$$H_A : \frac{10}{40} \neq \frac{5}{61}.$$

Step 3: Complete the table

Number without death in each period:

$$b_1 = 40 - 10 = 30, \qquad b_2 = 61 - 5 = 56.$$

Thus:

$$A = 10 + 5 = 15, \qquad B = 30 + 56 = 86, \qquad N = 101.$$

Step 4: Use Yates-corrected Chi-square

For a 2×2 table, the Yates-corrected statistic is

$$\chi^2 = \frac{\left(|a_1 n_2 - a_2 n_1| - \frac{N}{2}\right)^2 N}{A B n_1 n_2}.$$

So,

$$\chi^2 = \frac{\left(|10(61) - 5(40)| - \frac{101}{2}\right)^2 (101)}{(15)(86)(40)(61)} \approx 4.15.$$

Step 5: Decision

Since

$$4.15 > 3.841,$$

we reject H_0 at $\alpha = 0.05$.

Final interpretation

There is statistically significant evidence at the 5% level that the mortality rate was different between the two time periods. The observed data suggest that mortality decreased in the later period.

7. Example 2: Comparing Infection Rates Between Two Hospital Wards

Suppose a hospital compares infection rates in two wards.

	Infection	No infection	Total
Ward A	12	68	80
Ward B	5	65	70
Total	17	133	150

Observed rates

$$r_1 = \frac{12}{80} = 0.15 = 15\%$$
$$r_2 = \frac{5}{70} \approx 0.0714 = 7.14\%.$$

Expected counts

$$E_{11} = \frac{(80)(17)}{150} \approx 9.07, \qquad E_{12} = \frac{(80)(133)}{150} \approx 70.93,$$
$$E_{21} = \frac{(70)(17)}{150} \approx 7.93, \qquad E_{22} = \frac{(70)(133)}{150} \approx 62.07.$$

All expected counts are greater than 5, so the ordinary Chi-square test is acceptable.

Interpretation without full calculation

Since the infection rate in Ward A appears higher than in Ward B, the statistical test checks whether this difference is large enough to be attributed to chance alone.

If the p-value is less than 0.05, we conclude that the infection rates differ significantly. If the p-value is greater than 0.05, we conclude that the observed difference may be due to random variation.

8. Example 3: Small Sample Comparison — Use Fisher's Exact Test

Suppose a pilot study compares complication rates in two treatments.

	Complication	No complication	Total
Treatment A	1	9	10
Treatment B	5	5	10
Total	6	14	20

Observed complication rates:

$$\frac{1}{10} = 10\%, \qquad \frac{5}{10} = 50\%.$$

However, because the sample is very small and one cell is only 1, the Chi-square approximation is not reliable.

Therefore, Fisher's Exact Test is the correct choice.

Interpretation

Even if the percentages look very different, we must still use the appropriate statistical test before claiming a significant difference.

This is an important lesson in biomedical research:

A large numerical difference is not always a statistically significant difference.

9. How to Interpret the Result Correctly

When comparing rates, students often make one of two mistakes:

- They only compare the percentages and ignore statistical significance.
- They say "there is no difference" when the correct statement is "there is no statistically significant evidence of a difference."

Correct interpretation if p-value < 0.05

There is statistically significant evidence at the 5% level that the event rates differ between the two groups.

Correct interpretation if p-value > 0.05

There is not enough statistical evidence at the 5% level to conclude that the event rates differ between the two groups.

Very important

Failure to reject H_0 does not prove that the rates are exactly equal. It only means the data do not provide strong enough evidence of a difference.

10. Summary of the Lecture

- Comparing two rates means comparing

$$\frac{a_1}{n_1} \quad \text{and} \quad \frac{a_2}{n_2}.$$

- The problem can be written as a 2×2 table.
- Chi-square is used for sufficiently large samples.
- Yates correction is used for 2×2 tables when approximation needs improvement.
- Fisher's Exact Test is used for small samples or very small expected counts.
- The final conclusion must be written in statistical and biomedical language.

Homework (HW)

HW1

In a study of postoperative infection:

	Infection	No infection	Total
Male patients	18	82	100
Female patients	10	90	100
Total	28	172	200

Required:

1. Compute the infection rate in each group.
2. State H_0 and H_A.
3. Decide whether Chi-square is appropriate.
4. Write the biomedical interpretation.

HW2

A small pilot study gives the following table:

	Improved	Not improved	Total
Drug A	2	8	10
Drug B	6	4	10
Total	8	12	20

Required:

1. Compute the improvement rate in each group.
2. Decide whether Chi-square or Fisher is more appropriate.
3. Explain why.
4. Write the null and alternative hypotheses.

HW3

The following table shows mortality among infected patients in two hospitals:

	Death	Survival	Total
Hospital A	9	31	40
Hospital B	4	46	50
Total	13	77	90

Required:

1. Compute the mortality rate in each hospital.
2. Which hospital has the higher observed rate?
3. Which test would you use?
4. Write a complete interpretation.