Lecture 10

Hypothesis Testing: Comparing Proportions and Incidence Rates Between Two Populations

Section A: (Hypothesis tests) Comparing Proportions Between Two Populations: The "z-test" approach

2

Learning Objectives

- Upon completion of this lecture section you will be able to:
  - Estimate and interpret a p-value for comparing proportions between two populations using the "two sample z-test" approach
  - Explain why, even though there are three different measures of association (RD, RR, OR), only one p-value is needed

3

Example 1[1]

- Summary of response (y/n) by baseline CD4 count (≥ 250 vs. < 250)

	CD4 <250	CD4 ≥ 250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- Start with sample proportions:

$$\hat{P}_{CD\,4<250} = \frac{127}{503} = 0.253 \approx 0.25 \ (25\%)$$

$$\hat{P}_{CD\,4\geq 250} = \frac{79}{497} = 0.159 \approx 0.16 \ (16\%)$$

[1] http://inclass.kaggle.com/

4

Example 1: All Three CIs

- All three estimates and CIs

Risk Difference 0.09 (0.04, 0.14)
Relative Risk 1.56 (1.20, 2.01)
Odds Ratio 1.75 (1.27, 2.41)

5

Example 1: Two Sample "z-test"

- The two sample z-test is analogous to the two-sample t-test for comparing means of continuous data
- The approach is exactly the same, with different inputs
  - Specify the two competing hypotheses, null and alternative
  - Assume the null to be the truth
  - Compute how far the sample estimate is from the expected difference under the null assumption
  - Translate distance into a p-value
  - Make a decision

6

Example 1: Two Sample "z-test"

- Competing Hypotheses

$$H_o : p_{CD4<250} = p_{CD4\geq 250}$$
$$H_A : p_{CD4<250} \neq p_{CD4\geq 250}$$

- Notice, these competing hypothesis can also be presented as:

7

Example 1: Two Sample "z-test"

- Competing Hypotheses

$$H_o : p_{CD4<250} - p_{CD4\geq 250} = 0$$
$$H_A : p_{CD4<250} - p_{CD4\geq 250} \neq 0$$

- Compute the distance between the observed results and the expected (assuming the null) in standard errors

$$\frac{\hat{p}_{CD\ 4<250} - \hat{p}_{CD\ 4\geq 250}}{S\hat{E}\left(\hat{p}_{CD\ 4<250} - \hat{p}_{CD\ 4\geq 250}\right)} = \frac{0.25 - 0.16}{\sqrt{\frac{(.25)(0.75)}{503} + \frac{(0.17)(0.83)}{497}}} = \frac{0.09}{0.025} \approx 3.6$$

8

Example 1: Two Sample "z-test"

- Convert to a p-value
  - Our results was 3.6 standard errors above what is expected under the null (0). How likely is it to get a result 3.6 or more standard errors away from 0 just by chance when the null is true?

9

Example 1: Two Sample "z-test"

- Make a decision: the p-value is 0.0003

Interpretation: If the data came from two populations with the same proportion (probability) of responders, the chances of seeing the sample results (or something more extreme) are 0.0003 (3 in 10,000)

10

Example 2: Maternal/Infant HIV Transmission[2]

- Results

Results. From April 1991 through December 20, 1993, the cutoff date for the first interim analysis of efficacy, 477 pregnant women were enrolled; during the study period, 409 gave birth to 415 live-born infants. HIV-infection status was known for 363 births (180 in the zidovudine group and 183 in the placebo group). Thirteen infants in the zidovudine group and 40 in the placebo group were HIV-infected.

(at 18 mos)	AZT	Placebo	
HIV+	13	40	53
HIV-	167	143	310
	180	183	363

11

Example 3

- All three estimates and CIs

Risk Difference -0.15 (-0.22, -0.08)
Relative Risk 0.32 (0.18, 0.59)
Odds Ratios 0.27 (0.14, 0.53)

12

Example 2: Two Sample "z-test"

- Competing Hypotheses

$$H_o : p_{AZT} - p_{placebo} = 0$$
$$H_A : p_{AZT} - p_{placebo} \neq 0$$

- Compute the distance between the observed results and the expected (assuming the null) in standard errors

$$\frac{\hat{p}_{AZT} - \hat{p}_{placebo}}{SE(\hat{p}_{AZT} - \hat{p}_{placebo})} = \frac{0.07 - 0.22}{\sqrt{\frac{(.07)(0.93)}{180} + \frac{(0.22)(0.78)}{183}}} = \frac{-0.15}{0.036} \approx -4.17$$

13

Example 2: Two Sample "z-test"

- Convert to a p-value
  - Our results was 4.17 standard errors below what is expected under the null (0). How likely is it to get a result 4.17 or more standard errors away from 0 just by chance when the null is true?

14

Example 2: Two Sample "z-test"

- Make a decision: the p-value is (about) 0.0001

  Interpretation: If the data came from two populations with the same proportion (probability) of HIV transmission, the chances of seeing the sample results (or something more extreme) is 0.0001 (1 in 10,000)

15

Example 3: Aspirin and CVD: Women[3]

- From Abstract

BACKGROUND

Randomized trials have shown that low-dose aspirin decreases the risk of a first myocardial infarction in men, with little effect on the risk of ischemic stroke. There are few similar data in women.

METHODS

We randomly assigned 39,876 initially healthy women 45 years of age or older to receive 100 mg of aspirin on alternate days or placebo and then monitored them for 10 years for a first major cardiovascular event (i.e., nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes).

RESULTS

During follow-up, 477 major cardiovascular events were confirmed in the aspirin group, as compared with 522 in the placebo group

[3] Ridker P, et al. A Randomized Trial of Low-Dose Aspirin in the Primary Prevention of Cardiovascular Disease in Women. *New England Journal of Medicine* (2005). 352(13); 1293-1304

16

Example 3

- 2X2 of results

	Aspirin	Placebo	
CVD	477	522	999
No CVD	19,457	19,420	38,887
	19,934	19,942	39,876

17

Example 3

- All three estimates and CIs

  Risk Difference -0.002 (-0.005, 0.0008)
  Relative Risk 0.92 (0.80, 1.03)
  Odds Ratios 0.92 (0.80, 1.03)

- p-values from two-sample z-test ≈ 0.15
  - If the data came from two populations with the sample proportion of CVD cases, the chances of seeing the sample results (or something more extreme) is 0.15 (15 in 100)

18

Example 3

- Results

RESULTS

During follow-up, 477 major cardiovascular events were confirmed in the aspirin group, as compared with 522 in the placebo group, for a nonsignificant reduction in risk with aspirin of 9 percent (relative risk, 0.91; 95 percent confidence interval, 0.80 to 1.03;

19

Example 4

- HRT and Risk of CHD

Proportion of Women Developing CHD (Incidence)

	HRT	Placebo	
CHD	163	122	285
No CHD	8,345	7,980	16,325
	8,508	8,102	16,610

$$\hat{p}_{HRT} = \frac{163}{8,508} = 0.019 \quad (1.9\%)$$

$$\hat{p}_{Placebo} = \frac{122}{8,102} = 0.015 \quad (1.5\%)$$

Example 4

- All three estimates and CIs

  Risk Difference 0.004 (0.0001, 0.008)
  Relative Risk 1.27 (1.01, 1.60)
  Odds Ratios 1.28 (1.01, 1.62)

- p-values from two-sample z-test ≈ 0.042
  - If the data came from two populations with the same proportion (probability) who developed CHD, the chances of seeing the sample results (or something more extreme) is 0.042 (42 in 1,000)

21

Summary

- The "two sample z-test" provides a method for getting a p-value for testing two competing hypotheses about the true proportions of a binary outcome between two populations

$$H_o : p_1 = p_2$$
$$H_A : p_1 \neq p_2$$

22

Summary

- The two competing hypotheses can be expressed in terms of multiple measures of association

$$H_o : p_1 - p_2 = 0 \qquad H_o : \frac{p_1}{p_2} = 1 \qquad H_o : \frac{\left(p_1 / (1-p_1)\right)}{\left(p_2 / (1-p_2)\right)} = 1$$

$$H_A : p_1 - p_2 \neq 0 \qquad H_A : \frac{p_1}{p_2} \neq 1 \qquad H_A : \frac{\left(p_1 / (1-p_1)\right)}{\left(p_2 / (1-p_2)\right)} \neq 1$$

- As such, only one p-value is needed

23

Summary

- The test is performed using the observed risk difference ($\hat{p}_1 - \hat{p}_2$) and its estimate standard error , $\hat{SE}(\hat{p}_1 - \hat{p}_2)$

24

Section B: (Hypothesis tests) Comparing Proportions Between Two Populations: Chi-Squared and Fisher's Exact Tests

25

Learning Objectives

- Upon completion of this lecture section you will be able to:
  - Explain that the chi-square test for comparing proportions between two populations gives the exact same results as the "two sample z-test"
  - Explain the general principle of the chi-square approach
  - Interpret the results from an exact test for comparing proportions between two populations, Fisher's Exact Test
  - Explain the general principle of the Fisher's Exact Test approach
  - Name situations where Fisher's Exact Test is preferable to the chi-square/two sample z-test approach(es)

26

Example 1[1]

- Summary of response (y/n) by baseline CD4 count (≥ 250 vs. < 250)

	CD4 <250	CD4 ≥ 250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- Start with sample proportions:

$$\hat{P}_{CD\,4<250} = \frac{127}{503} = 0.253 \approx 0.25 \ (25\%)$$

$$\hat{P}_{CD\,4\geq 250} = \frac{79}{497} = 0.159 \approx 0.16 \ (16\%)$$

[1] http://inclass.kaggle.com/

27

Example 1: All Three CIs

- All three estimates and CIs

Risk Difference 0.09 (0.04, 0.14)
Relative Risk 1.56 (1.20, 2.01)
Odds Ratio 1.75 (1.27, 2.41)

p-value from "two-sample z-test" 0.0003

28

Example 1: Chi-Square(d) Approach

- The chi square test is a general test for comparing binary (or categorical) outcomes across two of more populations
- In the specific case of two proportions being compared across two populations, the results from the chi-square test and the two-sample z-test are identical: both depend on the same CLT based result
- The chi-square test can be extended to more comparisons in one test

29

Example 1: Two-Sample Chi Square Test

- The approach is exactly the same, with different inputs
  - Specify the two competing hypotheses, null and alternative
  - Assume the null to be the truth
  - Compute how far the samples based estimate is from what is expected under the null
  - Translate distance into a p-value
  - Make a decision

30

Example 1: Two-Sample Chi Square Test

- Competing Hypotheses

$$H_o : p_{CD4<250} = p_{CD4\geq 250}$$
$$H_A : p_{CD4<250} \neq p_{CD4\geq 250}$$

- Notice, these competing hypothesis can also be presented as:

31

Example 1: Two-Sample Chi Square Test

- Observed response (y/n) by baseline CD4 count (≥ 250 vs. < 250)

	CD4 <250	CD4 ≥ 250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- To start, a complimentary 2X2 table is computed, which shows the four cell counts expected if the null hypothesis is true

32

Example 1: Two-Sample Chi Square Test

- Observed response (y/n) by baseline CD4 count (≥ 250 vs. < 250)

	CD4 <250	CD4 ≥250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- Rationale behind computing expected cell counts

33

Example 1: Two-Sample Chi Square Test

- Observed response (y/n) by baseline CD4 count (≥ 250 vs. < 250)

	CD4 <250	CD4 ≥ 250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- Expected results under null

	CD4 ≥250	CD4 ≥ 250	
Respond	103.6		206
Not Respond			794
	503	497	1,000

34

Example 1: Two-Sample Chi Square Test

- The standardized distance between what was observed in the two samples, and what would be expected under the null is measured by the following formula:

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

- For this study: $\chi^2 = 13.37$

35

Example 1: Two-Sample Chi Square Test

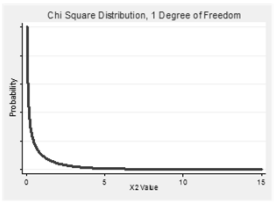
- The resulting difference is compared to the sampling distribution for all such distances occurring under the null hypothesis: this distribution is call a chi-square with 1 degree of freedom (χ^2_1) to get a p-value

- Idea behind one degree of freedom:

36

Example 1: Two-Sample Chi Square Test

- Chi square with one degree of freedom and p-value



37

Relationship

- Mathematical equivalence between the "two-sample z-test" and chi square test with 1 df

38

Example 1: Fisher's Exact

- Another approach for getting a p-value when comparing proportions between two populations is called Fisher's exact test
  - Can be used regardless of sample size
  - Generally gives same results as two-sample z/ chi square
  - Is computationally intensive (but not a problem for today's computers)

39

Example 1: Fisher's Exact

- Recall the results of this study

	CD4 <250	CD4 ≥ 250	
Respond	127	79	206
Not Respond	376	418	794
	503	497	1,000

- In the entire group there are 1,000 persons, and 206 responded to therapy

40

Example 1: Fisher's Exact

- Think of "marbles in a jar"

41

Example 1: Fisher's Exact

- Results for response to therapy/baseline CD4 count study
  - p-value from "two-sample z-test" 0.0003
  - p-value from chi square test 0.0003
  - p-value from Fisher's exact test 0.0003

Here, all 3 agree exactly: in smaller samples the p-values may differ slightly between the chi-square and Fisher's tests

42

Example 2: Maternal/Infant HIV Transmission[2]

- Results

Results. From April 1991 through December 20, 1993, the cutoff date for the first interim analysis of efficacy, 477 pregnant women were enrolled; during the study period, 409 gave birth to 415 live-born infants. HIV-infection status was known for 363 births (180 in the zidovudine group and 183 in the placebo group). Thirteen infants in the zidovudine group and 40 in the placebo group were HIV-infected.

(at 18 mos)	AZT	Placebo	
HIV+	13	40	53
HIV-	167	143	310
	180	183	363

43

Example 2

- All three estimates and 95% CIs

Risk Difference -0.15 (-0.22, -0.08)
Relative Risk 0.32 (0.18, 0.59)
Odds Ratios 0.27 (0.14, 0.53)

- Resulting p-values

two sample z-test / chi square 0.0001
Fisher's exact 0.0001

44

Example 2

- For both approaches, p-value is the same

Interpretation: If the data came from two populations with the same probability of HIV transmission, the chances of seeing the sample results (or something more extreme) is 0.0101 (1 in 10,000)

45

Example 3: Aspirin and CVD: Women[3]

- From Abstract

BACKGROUND
Randomized trials have shown that low-dose aspirin decreases the risk of a first myocardial infarction in men, with little effect on the risk of ischemic stroke. There are few similar data in women.

METHODS
We randomly assigned 39,876 initially healthy women 45 years of age or older to receive 100 mg of aspirin on alternate days or placebo and then monitored them for 10 years for a first major cardiovascular event (i.e., nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes).

RESULTS
During follow-up, 477 major cardiovascular events were confirmed in the aspirin group, as compared with 522 in the placebo group,

[3] Ridker P, et al. A Randomized Trial of Low-Dose Aspirin in the Primary Prevention of Cardiovascular Disease in Women. *New England Journal of Medicine* (2005). 352(13); 1293-1304

46

Example 3

- 2X2 of results

	Aspirin	Placebo	
CVD	477	522	999
No CVD	19,457	19,420	38,887
	19,934	19,942	39,876

47

Example 3

- All three estimates and CIs

Risk Difference -0.002 (-0.005, 0.0008)
Relative Risk 0.92 (0.80, 1.03)
Odds Ratios 0.92 (0.80, 1.03)

- p-values

two sample z-test/chi square 0.1511
Fisher's exact 0.1586

48

Example 4

- Sixty-five pregnant women, all who were classified as having a high risk of pregnancy induced hypertension, were recruited to participate in a study of the effects of aspirin on hypertension[1]
- The women were randomized to receive either 100 mg of aspirin daily, or a placebo during the third trimester of pregnancy

[1] Schiff, E. et al; The use of aspirin to prevent pregnancy-induced hypertension and lower the ratio of thromboxane A2 to prostacyclin in relatively high risk pregnancies, *New England Journal of Medicine* 321;6

49

Example 4

- Results in 2X2 table

Proportion of Women Developing Hypertension

	Aspirin	Placebo	
Hypertension	4	11	15
No Hypertension	30	20	50
	34	31	16,610

$$\hat{p}_{aspirin} = \frac{4}{34} = 0.12 \quad (12\%)$$

$$\hat{p}_{Placebo} = \frac{11}{31} = 0.35 \quad (35\%)$$

Example 4

- All three estimates and 95% CIs (aspirin compared to placebo)

  Risk Difference -0.24 (-0.44, -0.04)
  Relative Risk 0.33 (0.12, 0.93)
  Odds Ratio 0.24 (0.07, 0.83)

- p-values

  two sample z-test/chi-square 0.0234
  Fisher's Exact 0.0378

51

Summary

- The "two sample z-test" , chi-squared test and Fisher's Exact test all provide a method for getting a p-value for testing two competing hypotheses about the true proportions of a binary outcome between two populations

$$H_o : p_1 = p_2$$
$$H_A : p_1 \neq p_2$$

52

Summary

- The two competing hypotheses can be expressed in terms of multiple measures of association

$$H_o : p_1 - p_2 = 0 \qquad H_o : \frac{p_1}{p_2} = 1 \qquad H_o : \frac{\left(p_1 / (1-p_1)\right)}{\left(p_2 / (1-p_2)\right)} = 1$$

$$H_A : p_1 - p_2 \neq 0 \qquad H_A : \frac{p_1}{p_2} \neq 1 \qquad H_A : \frac{\left(p_1 / (1-p_1)\right)}{\left(p_2 / (1-p_2)\right)} \neq 1$$

- As such, only one p-value is needed

53

Summary

- The two-sample z-test and chi square give exactly the same result: however, the chi-square is usually what is referred to in the literature
- Fisher's exact test is a computer based test – it's results will usually align with the other two tests, but the resulting p-values may differ in slightly smaller samples
- Any of the three are generally appropriate for comparing proportions between two populations, and the resulting p-values are interpreted the same way

54

Summary

- While the mechanics differ between the three tests, the basic approach is exactly the same

55

Section C: Hypothesis Tests for Comparing Incidence Rates Between Two Populations

56

Learning Objectives

- Upon completion of this lecture section you will be able to
  - Describe two approaches to getting a p-value for comparing incidence rates between two populations:

    The two-sample z-approach is based on comparing the incidence rates on the natural log scale

    The log-rank test compares the Kaplan-Meier curves for the two groups (and can be extended to compare more than two populations)

57

Example 1

- Mayo Clinic: Primary Biliary Cirrhosis (PBC treatment), randomized clinical trial[1]

  Primary Research Question: How does mortality (and hence) survival for PBC patients randomized to receive DPCA (D-Penicillamine) compare to survival for PBC patients randomized to received a placebo?

1 Dickson E, et al. Trial of Penicillamine in Advanced Primary Biliary Cirrhosis. *New England Journal of Medicine*. (1985) 312(16): 1011-1015

58

Example 1

- Incidence rates for DPCA and placebo groups

DPCA: 872.5 years of follow-up, 65 deaths,n=158

$$\hat{IR}_{DPCA} = \frac{E_{DPCA}}{T_{DPCA}} = \frac{65 \text{ deaths}}{872.5 \text{ years}} \approx 0.075 \text{ deaths/yea r}$$

Placebo: 842.5 years of follow-up, 60 deaths, n=154

$$\hat{IR}_{placebo} = \frac{E_{placebo}}{T_{placebo}} = \frac{60 \text{ deaths}}{842.5 \text{ years}} \approx 0.071 \text{ deaths/yea r}$$

59

Example 1

- Incidence Rate Ratio

$$\hat{IRR} = \frac{\hat{IR}_{DPCA}}{\hat{IR}_{placebo}} = \frac{0.075 \text{ deaths/yea r}}{0.071 \text{ deaths/yea r}} \approx 1.06$$

Intepretations:

- The risk of death in the DPCA group (in the study follow-up period) is 1.06 times the risk in the placebo group

- Subjects in the DPCA group had 6% higher risk of death in the follow-up period when compared to the subjects in the placebo group

60

Example 1

- 95% CI for the IRR: (0.74, 1.52)
  - Notice that this 95% CI contains the null value of 1- what does this mean about the p-value for comparing the incidence rates?
- There are two approaches to getting a p-value, that yield very similar results
  - A "two sample z-test" approach
  - The log rank test

61

Example 1: two sample z-test

- Two sample z-test

Competing Hypotheses

$H_o: IR_{DPCA} = IR_{placebo}$ $\quad H_o: IRR_{DPCA/placebo} = 1$ $\quad H_o: \ln(IRR_{DPCA/placebo}) = 0$

$H_A: IR_{DPCA} \neq IR_{placebo}$ $\quad H_A: IRR_{DPCA/placebo} \neq 1$ $\quad H_A: \ln(IRR_{DPCA/placebo}) \neq 0$

- Approach
  - Assume H_o is true
  - Measure distance between $\ln(\hat{IRR})$ and 0 in standard errors
  - Convert to p-value and make decision

62

Example 1: two sample z-test

- Two sample z-test
  - Estimate standardized distance

$$z = \frac{\ln(\hat{IR})}{\hat{SE}(\ln(\hat{IR}))} = \frac{\ln(\hat{IR})}{\sqrt{\frac{1}{E_1} + \frac{1}{E_2}}} = \frac{0.06}{\sqrt{\frac{1}{60} + \frac{1}{65}}} = \frac{0.06}{0.18}$$

63

Example 1: two sample z-test

- Get a p-value

64

Example 1: two sample z-test

- Make a decision

The resulting p-value is 0.74. The result is not statistically significant and the decision is to "fail to reject the null"

65

Example 1: Log Rank Test

- Log Rank Test: this test compares the distance between the Kaplan Meier Curves in two samples to get a p-value

$H_o: IR_{DPCA} = IR_{placebo}$ $\quad H_o: S(t)_{DPCA} = S(t)_{placebo}$

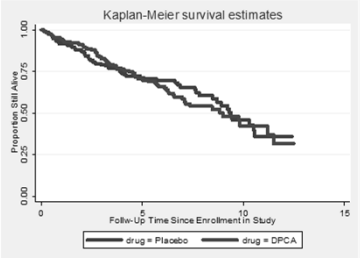
$H_A: IR_{DPCA} \neq IR_{placebo}$ $\quad H_A: S(t)_{DPCA} \neq S(t)_{placebo}$

- Idea: the log rank test compares the number of events observed at each event time in the two groups, to the number of expected events in each group
  - The discrepancies between the observed and expected event counts are aggregated across all event times and standardized by the uncertainty from sampling variability (standard error)

66

Example 1: Log Rank Test

- Kaplan-Meier Curves



67

Example 1: Log Rank Test

- The total, aggregated discrepancy, or distance between what is observed in the samples is compared to the distribution of such discrepancies across samples of the same size, when the null is true
  - This gets translated into a p-value
- For the DPCA/placebo comparison, the p-value from the log rank test is 0.75 : almost identical to the p-value from the two sample z approach

68

Example 2

- ART and Partner to Partner HIV Transmission[2]

RESULTS

As of February 21, 2011, a total of 39 HIV-1 transmissions were observed (incidence rate, 1.2 per 100 person-years; 95% confidence interval [CI], 0.9 to 1.7); of these, 28 were virologically linked to the infected partner (incidence rate, 0.9 per 100 person-years, 95% CI, 0.6 to 1.3). Of the 28 linked transmissions, only 1 occurred in the early-therapy group (hazard ratio, 0.04; 95% CI, 0.01 to 0.27; P<0.001). Subjects receiving early therapy had fewer treatment end points (hazard ratio, 0.59; 95% CI, 0.40 to 0.88; P=0.01).

2 Cohen M, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *New England Journal of Medicine*. (2011) 365(6): 493-505

69

Example 2

- ART and Partner to Partner HIV Transmission

"Of the 28 linked transmissions, only 1 occurred in the early therapy group (hazard ratio 0.04…)"

Note: hazard ratio and incidence rate ratio are (nearly) synonymous

So,

$$\widehat{IRR} = \frac{\widehat{IR}_{early}}{\widehat{IR}_{standard}} = \frac{\left(\frac{1 \text{ linked transmission}}{\text{total follow-up time, early therapy}}\right)}{\left(\frac{27 \text{ linked transmissions}}{\text{total follow-up time, standard therapy}}\right)} = 0.04$$

70

Example 2

- p-values

  two sample z-test 0.002
  log rank test: < 0.01 (as per authors)

- Decision

  reject the null hypothesis

71

Example 2

- Differences between the two sample z-test and the log rank

72

Example 2

- Interpretation (I will use the results I computed)

  In a study of 1,763 HIV sero-discordant couples, the risk of partner-to-partner transmission among the 866 randomized to receive early ART therapy was 96% lower than among the 877 randomized to receive standard ART therapy. (p=0.002) After accounting for sampling variability, the early ART therapy could reduce risk of partner transmission from 69% to 99% at the population level.

73

Example 2

- Interpretation (I will use the results I computed)

  The p-value of 0.002 means that if the underlying rate of partner to partner transmission were the same in the populations of sero-discordant couples given early or standard ART, then the chances of getting a sample incidence rate of 0.04 (or something more extreme is 2 out of 1000)

74

Example 3

- Maternal Vitamin Supplementation and Infant Mortality[3]

ABSTRACT
Background: The effect of vitamin A supplementation on the survival of infants aged <6 mo is unclear. Because most infant deaths occur in the first few month of life, maternal supplementation may improve infant survival.
Objectives: The objective was to assess the effect of maternal vitamin A or β-carotene supplementation on fetal loss and survival of infants <6 mo of age.
Design: Married women of reproductive age in 270 wards of Sarlahi district, Nepal, were eligible to participate. Wards were randomly assigned to have women receive weekly doses of 7000 μg retinol equivalents as retinyl palmitate (vitamin A), 42 mg *all-trans*-β-carotene, or placebo. Pregnancies were followed until miscarriage, stillbirth, maternal death, or live birth of one or more infants, who were followed through 24 wk of age.

3 Katz J, West K et al. Maternal low-dose vitamin A or B-carotene supplementation has no effect on fetal loss and early infant mortality: a randomized cluster trial in Nepal. *American Journal of Clinical Nutrition* (2000) Vol. 71, No. 6, 1570-1576.

75

Studies Involving Follow-Up Over Time: Example 3

- Incidence Rate Ratio: 3 have three groups, can make 1 the "reference" or comparison group: I suggest placebo as the reference group

$$\widehat{IR}_{vitA} = \frac{\hat{IR}_{vitA}}{\hat{IR}_{placebo}} = \frac{0.00041 \quad \text{deaths/day}}{0.00039 \quad \text{deaths/day}} \approx 1.05$$

$$\widehat{IR}_{BC} = \frac{\hat{IR}_{BC}}{\hat{IR}_{placebo}} = \frac{0.00039 \quad \text{deaths/day}}{0.00039 \quad \text{deaths/day}} \approx 1.00$$

76

Studies Involving Follow-Up Over Time: Example 3

- Incidence Rate Ratios with 95% CIS and p-values:

	two sample z	log rank
$\widehat{IR}_{vitA}$: 1.05 (0.87, 1.28)	0.55	0.52
$\widehat{IR}_{BC}$: 1.00 (0.84, 1.25)	0.84	0.82

77

Summary

- Both the two-sample z-test and the log rank can be used to test competing null and alternative hypothesis about time to event data
- The log rank is most commonly presented in the literature, but the two sample z-test is a nice, easy to implement by hand approach, that is very similar in its approach to the two-sample t-test for comparing means, and the two-sample z-test for comparing proportions

78

Summary

- Because of slightly different mechanics the p-values from both tests may differ slightly in value
- Both tests use the same logic as all other hypothesis tests we've seen

79

Section D: Debriefing on the p-value and Hypothesis Testing, Part 2

80

Two Sided Hypothesis Tests

- The way we've demonstrated performing hypothesis tests are for two sided hypothesis tests, and result in two-sided p-values
- "Two-sided" refers to the method of getting p-value that measures being as far or farther from the null value (as extreme of more extreme) in either direction

81

Two Sided Hypothesis Tests

- It is possible (and sometime more logical perhaps) to perform a one-sided test. For example..

82

Two Sided Hypothesis Tests

- However, one-sided tests are not commonly presented in the literature, and in fact tend to raise "suspicion"
  - The appropriate one sided test will always result in a p-value that is half that of the two-sided p-value

83

Keeping Track!

- We have named a lot of tests in lecture sets 9 and 10
- Regardless of the name, the approach is universally the same

84

Keeping Track!

- The names distinguish both the tests in terms of
  - The type of data being compared between the two groups
  - The specific mechanics of getting to a p-value
- One can always look up the name of the appropriate test(s) given the data being compared- the important thing to note is that all are conceptually the same, and the resulting p-values:
  - Have the same interpretation across the tests
  - Will agree with the corresponding confidence intervals for the chosen measure of association with regards to the null hypothesis

85

Keeping Track!

- You will undoubtedly see other tests in the literature that we have not yet covered or will not cover in this class
- Again, though, if you can figure out what is being compared via the test, then you can interpret the p-value in the context of the comparison
- In then next lecture set we will discussion extensions to the tests covered in lecture 9 and 10 to handle comparisons between more than 2 populations in one test

86

Keeping Track!

- Thus far, for comparing two populations:

Means	Proportions	Time-to-Event

87

Different Alpha Levels

- Just as it is possible to have difference levels for confidence intervals other than 95% (90%, 99% etc..) , it is possible to evaluate p-values at different rejection levels (α=0.10, α=0.01 etc..)

88

Different Alpha Levels

- However, the standard in research is to use 95% CIs and a rejection (alpha, type 1 error) level of 0.05

89