Comparing Proportions between Two Independent Populations

John McGready
Johns Hopkins University

Lecture Topics

- Using CIs for difference in proportions between two independent populations

- Large sample methods for comparing proportions between two populations
  - Normal method
  - Chi-squared test

- Fisher's exact test

- Relative risk

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

The Two Sample z-Test for Comparing Proportions between Two Independent Populations: The Confidence Interval Approach

Comparing Two Proportions

- We will motivate by using data from the Pediatric AIDS Clinical Trial Group (ACTG) Protocol 076 Study Group*

- Study design
  - "We conducted a randomized, double-blinded, placebo-controlled trial of the efficacy and safety of zidovudine (AZT) in reducing the risk of maternal-infant HIV transmission"
  - 363 HIV infected pregnant women were randomized to AZT or placebo

Notes: *Conner, E., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment, *New England Journal of Medicine* 331: 18.

Comparing Two Proportions

- Results
  - Of the 180 women randomized to AZT group, 13 gave birth to children who tested positive for HIV within 18 months of birth
  - Of the 183 women randomized to the placebo group, 40 gave birth to children who tested positive for HIV within 18 months of birth

Notes on Design

- Random assignment of Tx
  - Helps insure two groups are comparable
  - Patient and physician could not request particular treatment

- Double blind
  - Patient and physician did not know treatment assignment

Observed HIV Transmission Proportions

- AZT

$$\hat{p}_{AZT} = \frac{13}{180} = 0.07 = 7\%$$

- Placebo

$$\hat{p}_{PLAC} = \frac{40}{183} = 0.22 = 22\%$$

HIV Transmission Proportions: 95% CIs

```
. cii 180 13

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |        180    .0722222     .019294        .0390137    .1203358

. cii 183 40

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |        183    .2185792    .0305507         .160984    .2855248
```

Notes on HIV Transmission Proportions

- Is the difference significant, or can it be explained by chance?

- Since CIs do not overlap suggests significant difference
  - Can we compute a confidence interval on the difference in proportions?
  - Can we compute a p-value?

Sampling Distribution: Difference in Sample Proportions

- Since we have large samples we know the sampling distributions of the sample proportions in both groups are approximately normal

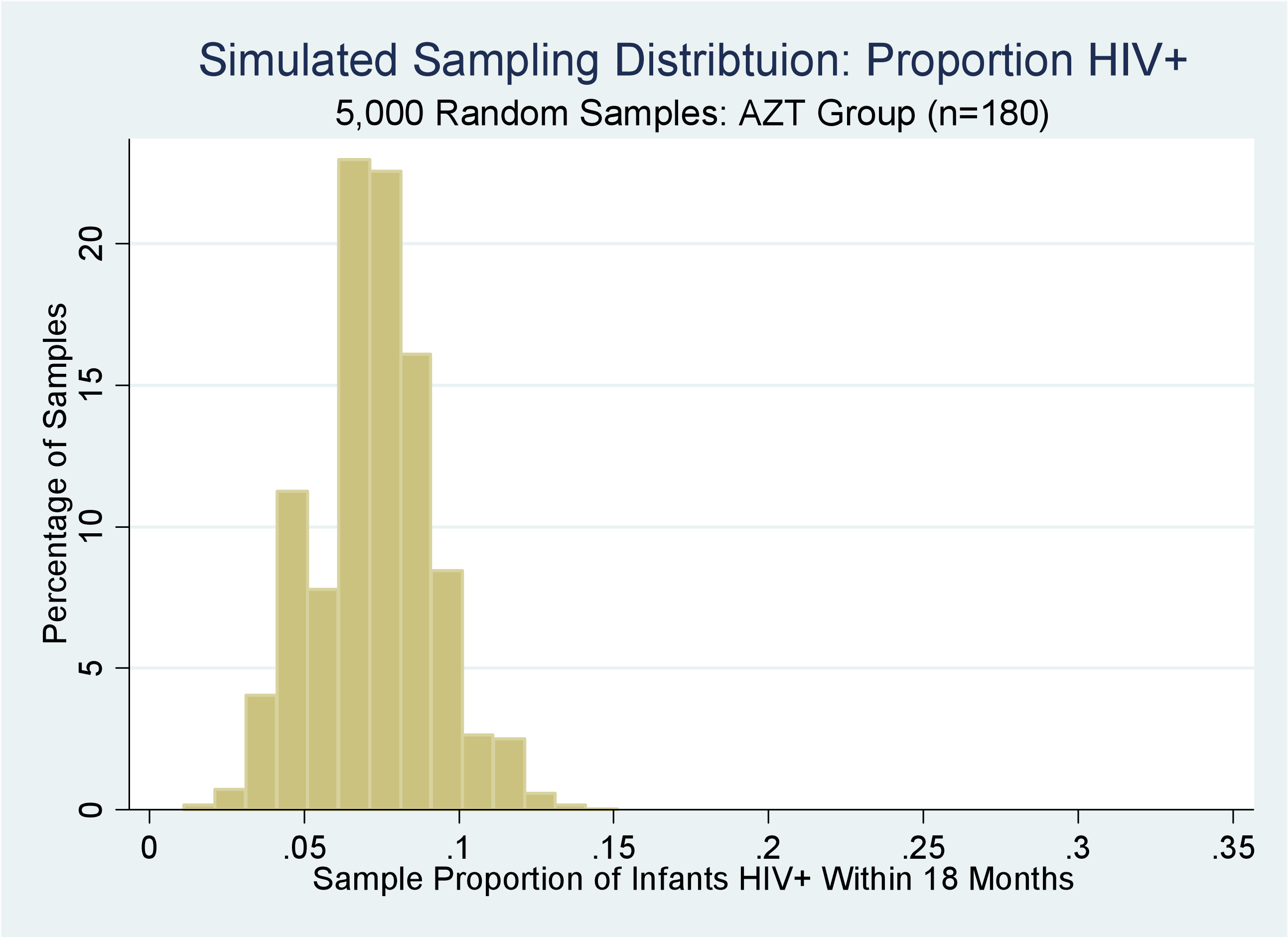
- It turns out the difference of quantities, which are (approximately) normally distributed, are also normally distributed

Sampling Distribution: Difference in Sample Proportions

- So, the big news is . . .
  - The sampling distribution of the difference of two sample proportions, each based on large samples, approximates a normal distribution
  - This sampling distribution is centered at the true (population) difference, $p_1 - p_2$

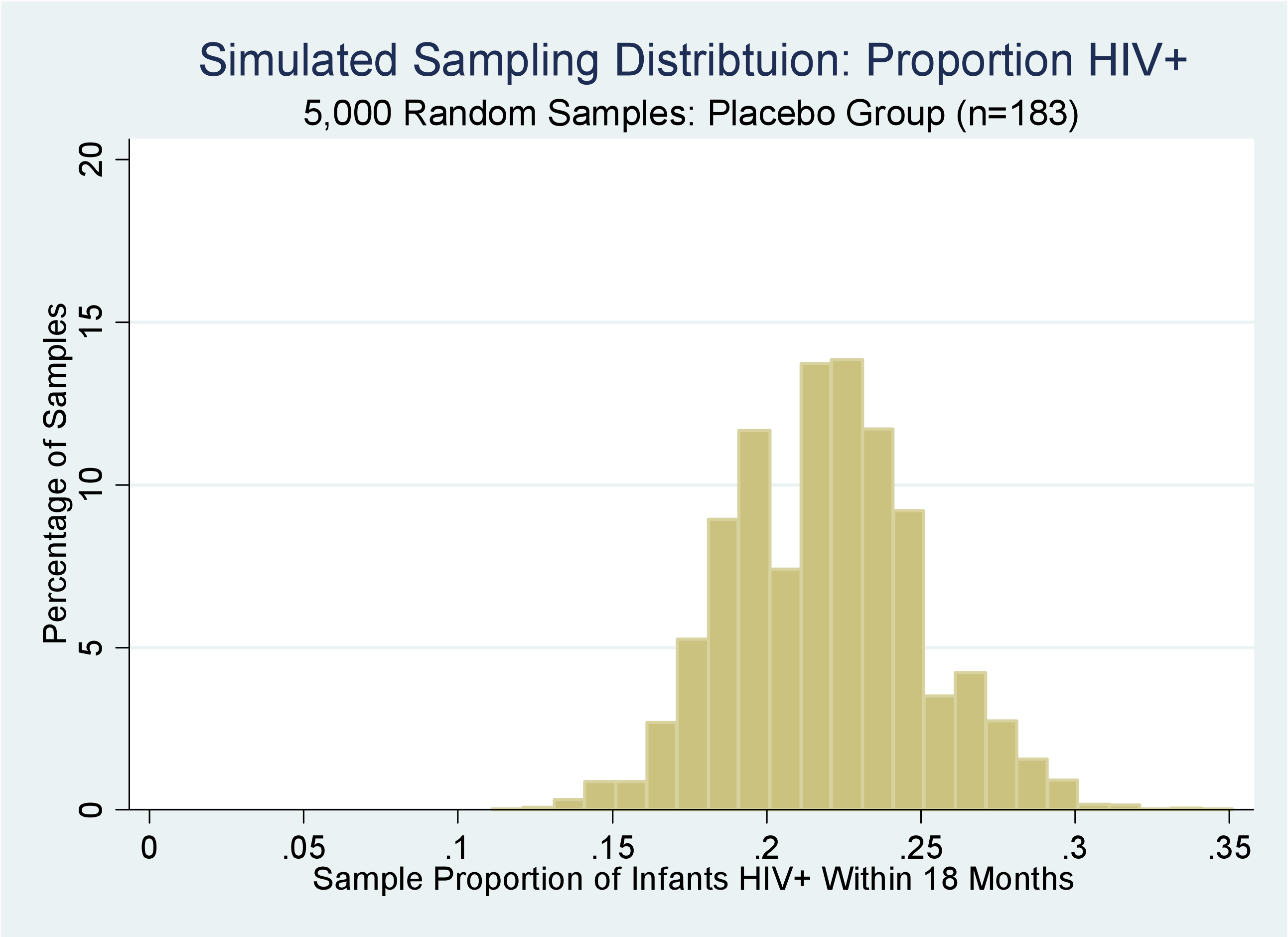
Sampling Distribution: AZT Group

- Simulated sampling distribution of sample proportion: AZT group



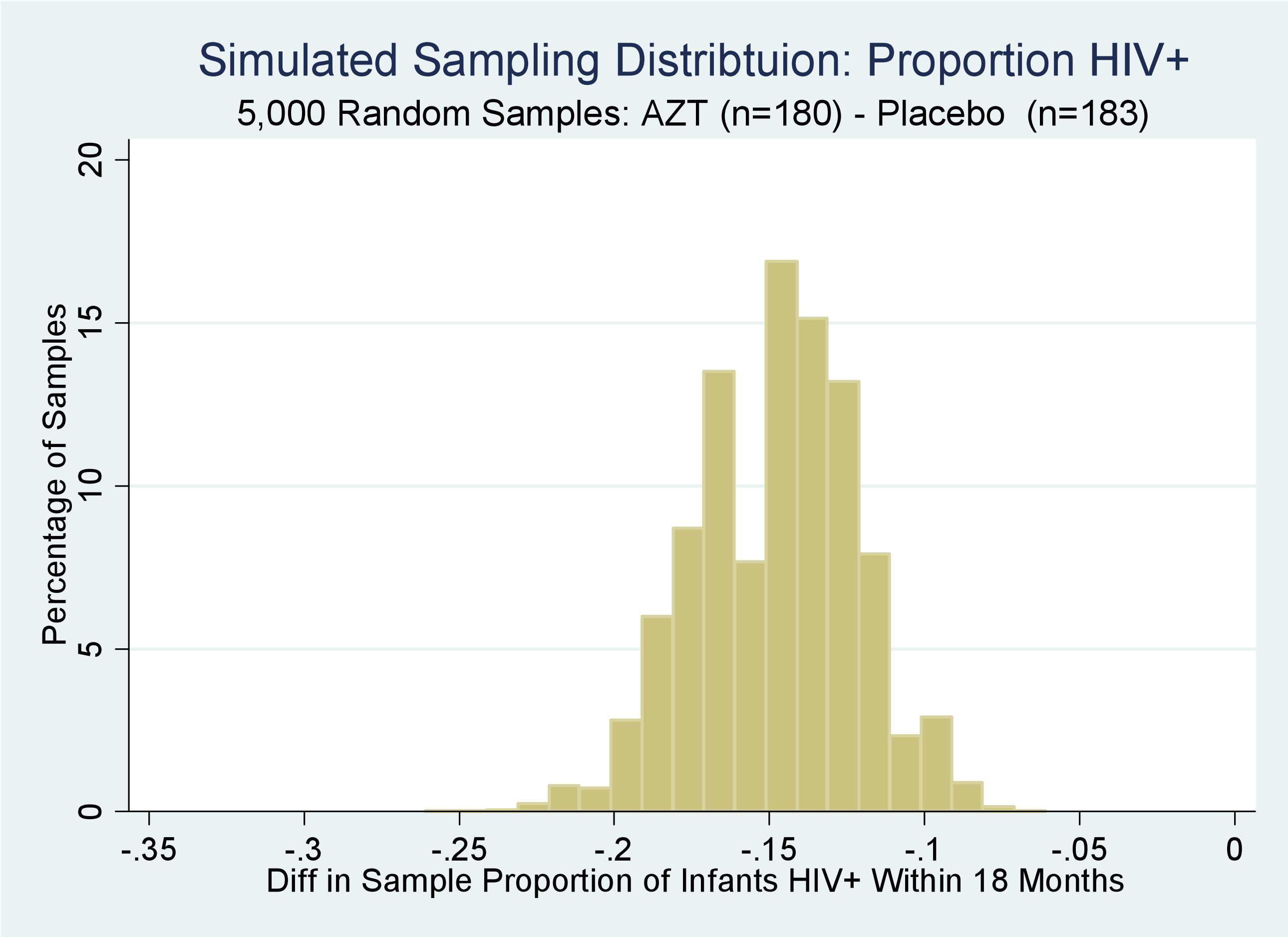
Sampling Distribution: Difference in Sample Proportions

- Simulated sampling distribution of sample proportion: placebo group



Sampling Distribution: Difference in Sample Proportions

- Simulated sampling distribution of difference in sample proportions: AZT - placebo



95% Confidence Interval for Difference in Proportions

- Our most general formula

$$best\ estimate\ from\ sample \pm 2 \times SE(best\ estimate\ from\ sample)$$

- The best estimate of a population difference based on sample proportions:

$$\hat{p}_1 - \hat{p}_2$$

- Here, $\hat{p}_1$ may represent the sample proportion of infants HIV positive (within 18 months of birth) for 180 infants in the AZT group, and $\hat{p}_2$ may represent the sample proportion of infants HIV positive (within 18 months of birth) for 183 infants in the AZT group

95% CI for Difference in Proportions: AZT Study

- So, $\hat{p}_1 - \hat{p}_2 = 0.07 - 0.22 = -0.15$: hence the formula for the 95 CI for p1 - p2 is:

$$-0.15 \pm 2 \times SE(\hat{p}_1 - \hat{p}_2)$$

- Where $SE(\hat{p}_1 - \hat{p}_2)$ = standard error of the difference of two sample proportions

Standard Error of Difference in Proportions

- Statisticians have developed formulas for the standard error of the difference
- These formulas depend on sample sizes in both groups and sample proportions in both groups
- The $SE(\hat{p}_1 - \hat{p}_2)$ is greater than either $SE(\hat{p}_1)$ or $SE(\hat{p}_2)$
  - Why do you think this is?

Principle

- Variation from independent sources can be added
  - Why do you think this is additive?

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 \times (1 - p_1)}{n_1} + \frac{p_2 \times (1 - p_2)}{n_2}}$$

- Of course, we don't know p_1 and p_2: so we estimate with $\hat{p}_1$ and $\hat{p}_2$ to get an estimated standard error:

$$\hat{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 \times (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \times (1 - \hat{p}_2)}{n_2}}$$

Comparing Two Independent Groups: HIV/AZT Study

- Recall the data from the Infant HIV/ AZT study

Group		
	AZT	Placebo
Number of subjects (n)	64	68
Proportion Infants HIV+ Within 18 Months	0.07	0.22

$$S\hat{E}\,(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.07 \times .93}{180} + \frac{.22 \times .78}{183}} \approx .36$$

95% CI for Difference in Proportions: HIV/ AZT Study

- So in this example, the estimated 95% for the true difference in proportions of infants contracting HIV between the AZT and placebo groups:

$$-0.15 \pm 2 \times S\hat{E}(\hat{p}_1 - \hat{p}_2)$$

$$-0.15 \pm 2 \times .036$$

$$-0.15 \pm .072$$

$$-0.222 \ \ to \ 0.078 \approx$$

$$-22\% \ to \ -8\%$$

Summary: AZT Study

- **Results**
  - The proportion of infants who tested positive for HIV within 18 months of birth was seven percent (95% CI 4 -12%) in the AZT group and twenty-two percent in the placebo group (95% CI 16 - 28%)
  - The study results estimate the absolute decrease in the proportion of HIV positive infants born to HIV positive mothers associated with AZT to be as low as 8% and as high as 22%