Cohort Studies

Zuber D. Mulla

1. **Systematic reviews, meta-analyses**
2. **Randomized controlled trials with definitive results**
3. **Randomized controlled trials with non-definitive results**
4. **Cohort studies**
5. **Case-control studies**
6. **Cross sectional surveys**
7. **Case reports**

A major type of observational study

- **A study in which subjects are classified on the basis of the presence or absence of exposure to a suspected risk factor for a disease or other outcome.**

Hennekens & Buring,
Epidemiology in Medicine

Cohort (Follow-Up) Studies

- **At the start of the study, all of the potential subjects must be free of the disease (the outcome).**

- **The two groups are compared to one another: risk or rate of the disease in the exposed vs. risk or rate of disease in the nonexposed**

Two Types of Cohort Studies

- **Retrospective**
- **Prospective**

Retrospective Cohort

- **Both the exposure and outcome have already occurred when the study is started.**
- **Start with the exposure**
- **Follow subjects up**

Example

- **In January of 2008, an OB/GYN contacts a large health maintenance (HMO) organization and identifies women who were free of ovarian cancer in 1985 who had a bilateral tubal ligation (BTL) in 1985 and women who free of ovarian cancer in 1985 who never had a BTL as of 1985 or afterwards.**

Retrospective Cohort

- **She then links the HMO records with various state cancer registries to determine which of the women who had a BTL in 1985 developed ovarian cancer by 2006 and which of the women who never had a BTL developed ovarian cancer by 2006.**

Retrospective Cohort

- **Retrospective: Because by the time the investigator started the study the exposure (BTL) and the outcome (ovarian cancer) had both occurred.**

- **Cohort: She started with exposure, not the outcome. Incidence.**

- **Temporality is intact. Cause → Effect**

Prospective Cohort

- **Start with your two groups: exposed and nonexposed.**
- **Follow them for minutes, hours, days, months, or years for one or more outcomes.**
- **Moving forward.**

Advantages

- **Incidence**
- **Temporality**
- **Multiple outcomes**
- **Valuable for studying rare exposures**

Disadvantages

- **If prospective, can be very expensive and time consuming.**
- **If retrospective, requires the availability of adequate records.**
- **Validity can be affected by losses to follow-up.**

Well-Known Prospective Cohort Studies

- **Framingham Heart Study**
- **Nurses' Health Study**

2 x 2 Table for a Cohort Study or Clinical Trial with Equal Lengths of Follow Up for All Subjects

	Had outcome	Didn't have outcome
Exposed	A	B
Unexposed	C	D

$$\text{Risk} = \frac{\text{Number of new cases of a disease during a given period of time}}{\text{Total population at risk}}$$

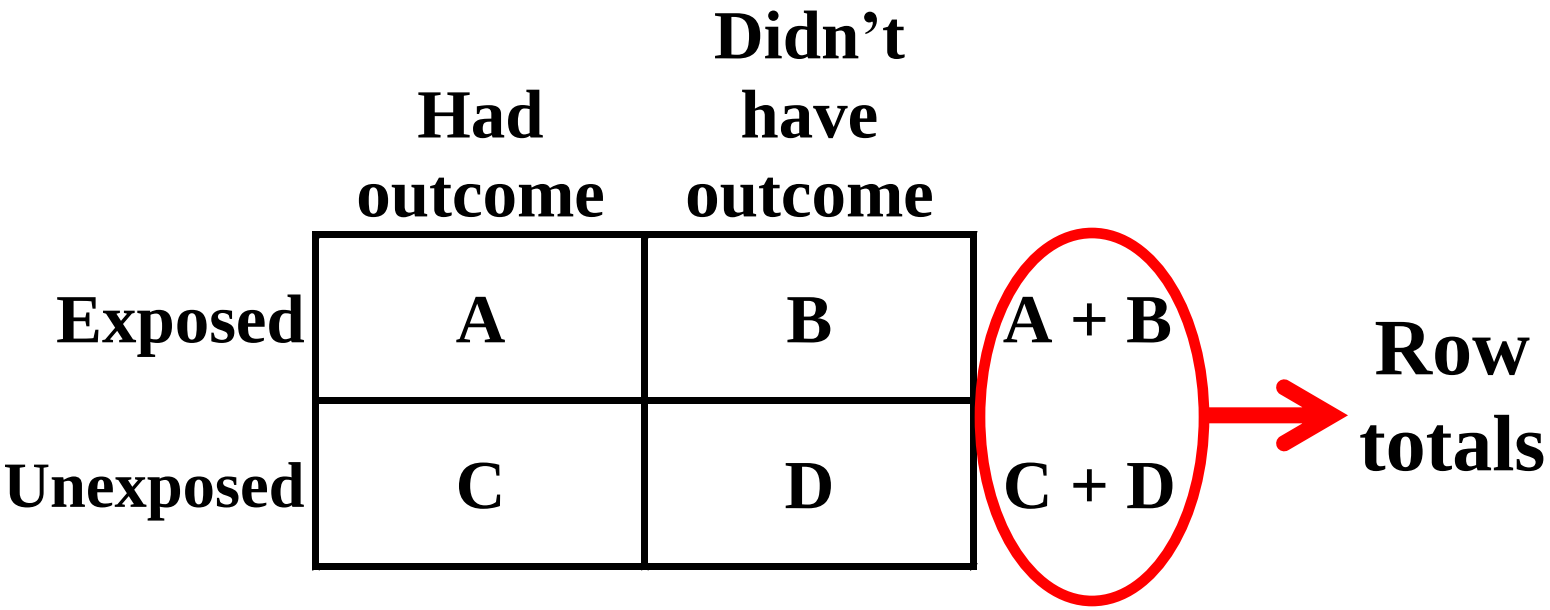
Synonyms for Risk

- **Cumulative incidence**
- **Incidence proportion**

Relative Risk (RR)

$$RR = \frac{\text{Incidence in the exposed}}{\text{Incidence in the unexposed}}$$

2 x 2 Table


Had outcome
Didn't have outcome
Exposed
A
B
A + B
Unexposed
C
D
C + D
Row totals

Relative Risk

Incidence in the exposed = A / (A + B)

Incidence in the unexposed = C / (C + D)

Relative Risk

$$= \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

Possible Values of RR

Interpretation

RR = 1.25

The exposed group was 1.25 times as likely or 25% more likely to develop the outcome than the unexposed group.

Interpretation

RR = 0.80

The exposed group had 80% of the probability of developing the outcome than the unexposed, or was 20% less likely than the unexposed to develop the outcome.

Null Value: No Association

RR = 1

Some Examples of Strong Associations

RR = 5.00

RR = 0.20

Some Examples of Weak Associations

RR = 1.04

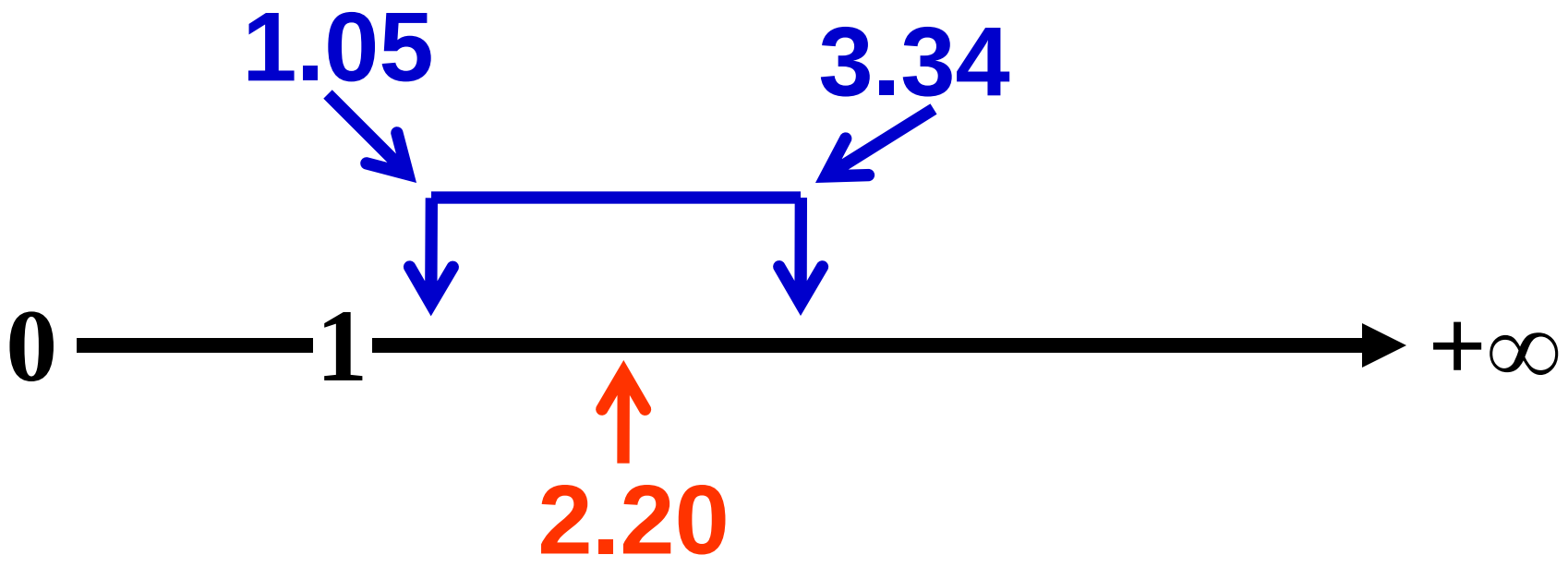
RR = 0.95

Statistical Significance: OR and RR

If P-value is 0.05 or less, then result is statistically significant

If the 95% CI excludes one (e.g., 1.05 – 3.34), then the result is statistically significant and the P-value will be 0.05 or less

Does the CI enclose 1?


1.05
3.34
0
1
+∞
2.20

Statistical Significance: OR and RR

If the 99% CI excludes one (e.g., 1.05 – 3.34), then the P-value will be 0.01 or less

Incidence Rate

- **Not a proportion so range is 0 to $+\infty$**

- **The momentary rate at which cases are occurring within a population**

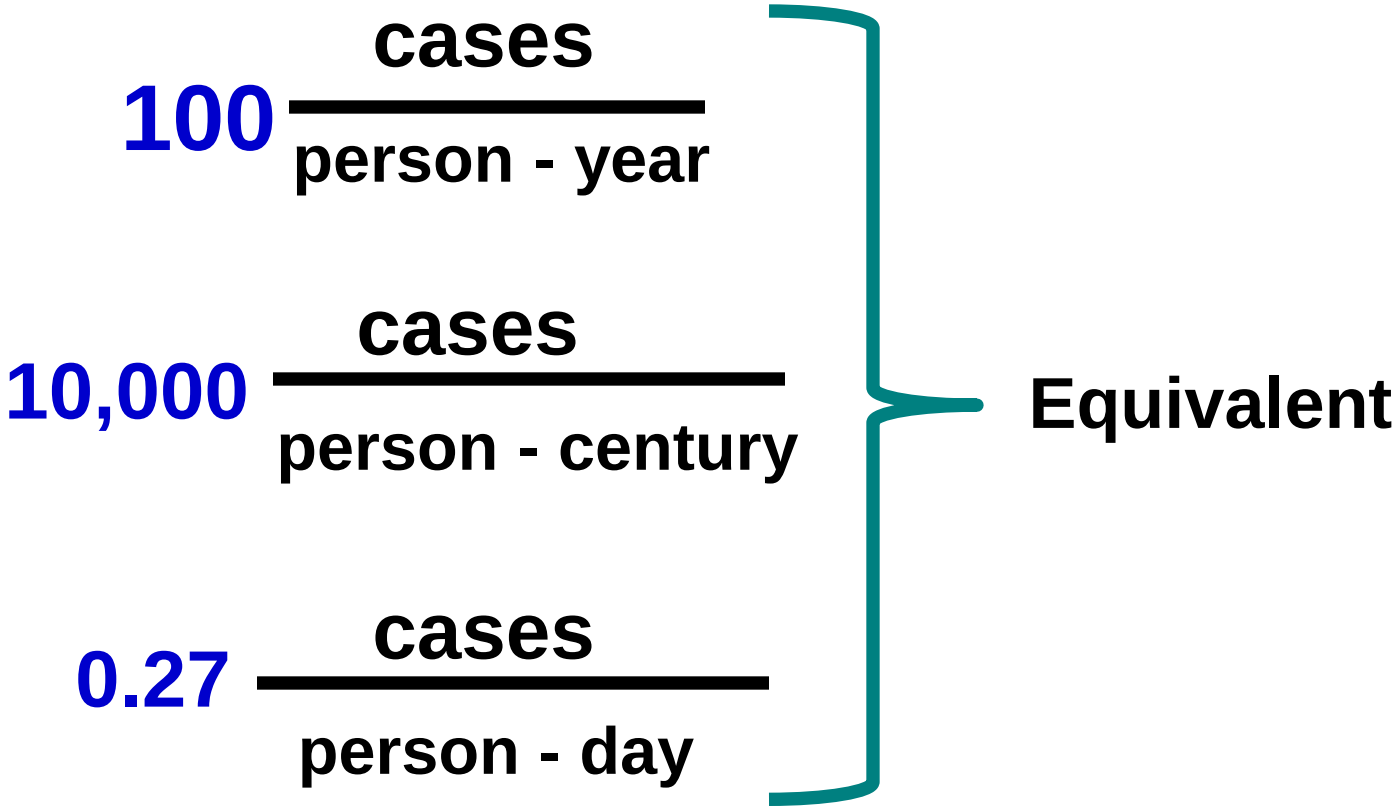
Incidence Rate, Unlike Incidence Risk, Can Be Greater than 1

$$100 \frac{\text{cases}}{\text{person - year}}$$

$$10{,}000 \frac{\text{cases}}{\text{person - century}}$$

$$0.27 \frac{\text{cases}}{\text{person - day}}$$

Incidence Rate, Unlike Incidence Risk, Can Be Greater than 1



$$\frac{10{,}000 \text{ cases}}{\text{person - century}} \times \frac{\text{person - century}}{36{,}525 \text{ person - day}} = \frac{\qquad}{\qquad}$$

$$\frac{10{,}000 \text{ cases}}{\text{person - century}} \times \frac{\text{person - century}}{36{,}525 \text{ person - day}} = \frac{0.27 \text{ cases}}{\text{person - day}}$$

Incidence Rate

- **Appropriate when there are varying periods of follow up**

- **Synonym: incidence density**

$$\text{Incidence rate} = \frac{\text{Number of new cases of a disease during a given period of time}}{\text{Total person-time of observation}}$$

Calculating Incidence Rates in the Exposed (or Treated) and in the Unexposed (or Untreated)

	Had outcome	Didn't have outcome	
Exposed	A	—	PT_e
Unexposed	C	—	PT_{ne}

Relative Rate (Incidence Density Ratio)

$$= \frac{\frac{A}{PT_e}}{\frac{C}{PT_{ne}}}$$

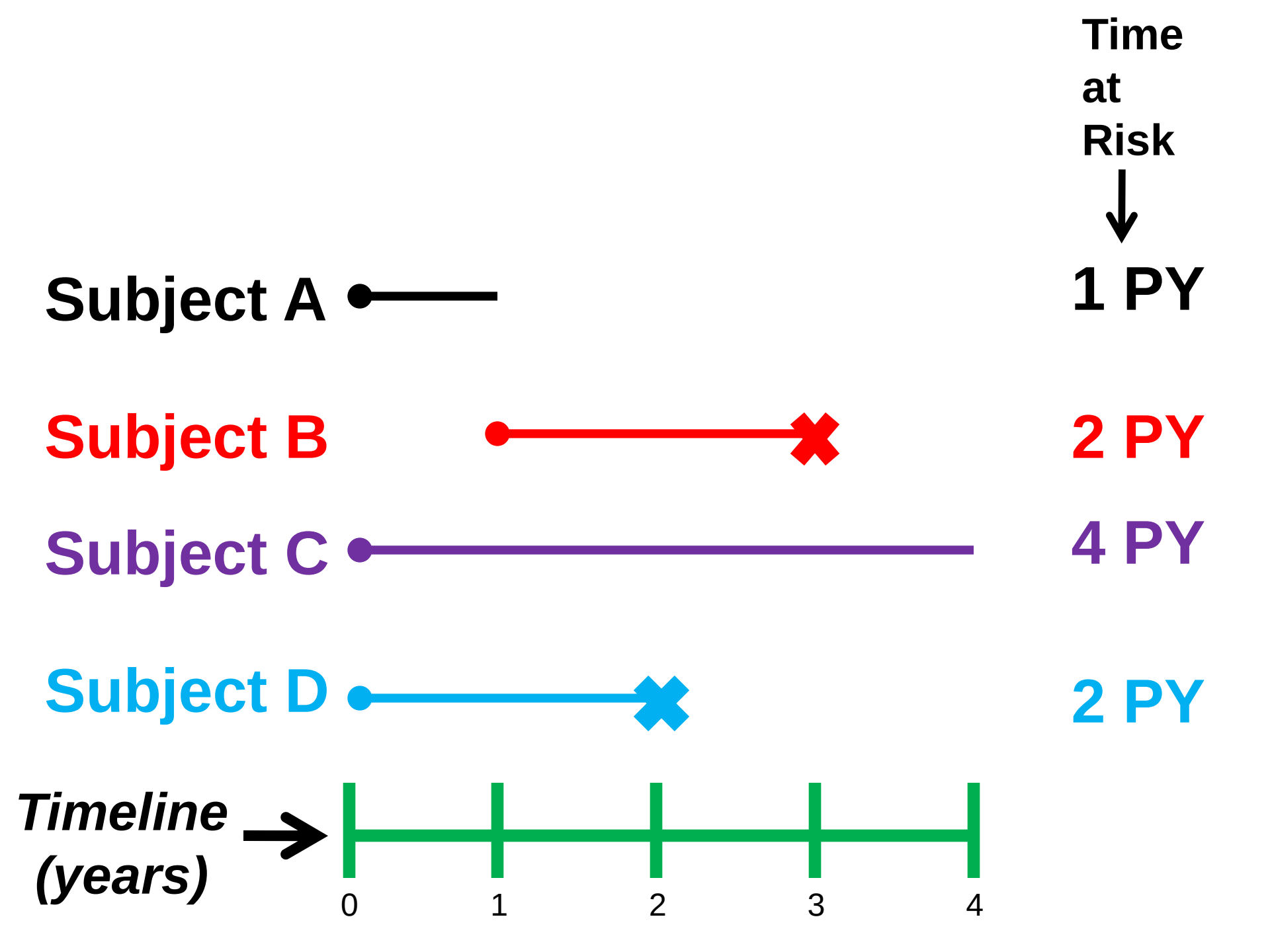
Calculation of Incidence Density in the Exposed Group:

Four subjects who enrolled in a four-year study at different times and were followed for varying lengths.

● = Start of follow-up

— = Time followed

✖ = Development of outcome


Time at Risk
Subject A
1 PY
Subject B
2 PY
Subject C
4 PY
Subject D
2 PY
Timeline (years)
0
1
2
3
4

Calculation of Incidence Density (ID) in the Exposed Group

- **Subject A dropped out of the four-year study before the study was completed. He did not have the outcome of interest when he dropped out.**

- **He was censored after one year of follow up. Subject C also censored.**

Calculation of Incidence Density (ID) in the Exposed Group

- 9 person-years of follow-up

ID = 2 cases divided by 9 person-years

ID = 0.222 / person-year or
2.22 / 10 person-years

Odds Ratios in Cohort Studies and Clinical Trials

- **Allowable, and very common due to the use of logistic regression**

- **Incidence odds ratio (rather than an exposure odds ratio or prevalence odds ratio)**

Try log-binomial before using logistic regression

- **Robbins AS, Chao SY, Fonseca VP. What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Ann Epidemiol* 2002;12: 452-4.**

- **Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005; 162: 199-200.**

Data Analysis of Cohort Studies & Clinical Trials

Technique or Statistical Model	Result
Standard life tables	Probability of death
Life tables with small intervals (product-limit or Kaplan-Meir)	Kaplan-Meier survival curves
Poisson regression	Rate ratio
Logistic regression	Odds ratio
Cox regression (proportional hazards)	Hazard ratio
Log-binomial regression	Risk ratio

Cited Reference

Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.