Measures of association: guide for data analysis

Gustaf Edgren, PhD
Karolinska Institutet

Outline

- Repetition of basic measures of association
- Step by step introduction to epidemiologic data analysis
- Application

The 2-by-2 table

Exposure \ Outcome	+	-	
+	a	b	a + b
-	c	d	c + d

Relative risks

$$RR = \frac{Risk_{\text{exposed}}}{Risk_{\text{unexposed}}}$$

Risk ratios

$$CIR = \frac{CI_{\text{exposed}}}{CI_{\text{unexposed}}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Rate ratios

$$IRR = \frac{IR_{\text{exposed}}}{IR_{\text{unexposed}}} = \frac{\frac{a}{pyrs_{\text{exposed}}}}{\frac{c}{pyrs_{\text{unexposed}}}}$$

Odds ratios

$$OR = \frac{odds_{\text{exposed}}}{odds_{\text{unexposed}}} = \frac{a/\text{b}}{\text{c}/d} = \frac{a \times d}{b \times c}$$

Absolute measures

$$RD = Risk_{\text{exposed}} - Risk_{\text{unexposed}}$$

Risk differences

$$CID = CI_{\text{exposed}} - CI_{\text{unexposed}} = \frac{a}{a+b} - \frac{c}{c+d}$$

$$IRD = IR_{\text{exp}} - IR_{\text{unexp}} = \frac{a}{pyrs_{\text{exposed}}} - \frac{c}{pyrs_{\text{unexposed}}}$$

Step-by-step guide for data analysis

1. Formulate what question you are trying to answer? Be specific:
   - What is the exposure?
   - What is the outcome?
2. Summarize the data
3. Calculate measures of disease occurrence according to exposure status
4. Calculate measure of association, comparing exposure groups
5. Interpret results

Example 1 – Storage time

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Duration of Red-Cell Storage and Complications after Cardiac Surgery

Colleen Gorman Koch, M.D., Liang Li, Ph.D., Daniel I. Sessler, M.D., Priscilla Figueroa, M.D., Gerald A. Hoeltge, M.D., Tomislav Mihaljevic, M.D., and Eugene H. Blackstone, M.D.

ABSTRACT

1. Formulate the question

METHODS

We examined data from patients given red-cell transfusions during coronary-artery bypass grafting, heart-valve surgery, or both between June 30, 1998, and January 30, 2006. A total of 2872 patients received 8802 units of blood that had been stored for 14 days or less ("newer blood"), and 3130 patients received 10,782 units of blood that had been stored for more than 14 days ("older blood"). Multivariable logistic regression with propensity-score methods was used to examine the effect of the duration of storage on outcomes. Survival was estimated by the Kaplan–Meier method and Blackstone's decomposition method.

- Is mortality higher among recipients of red-cell units stored for more than 14 days compared to recipients of units stored for less than 14 days?

1. Formulate the question, cont.

- What is the exposure?
  - Exposure to blood stored for more than 14 days
- What is the outcome?
  - Death within 2 years of first transfusion

2. Summarize the data

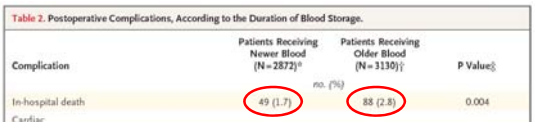

Table 2. Postoperative Complications, According to the Duration of Blood Storage.

Complication	Patients Receiving Newer Blood (N = 2872)*	Patients Receiving Older Blood (N = 3130)†	P Value‡
	no. (%)		
In-hospital death	49 (1.7)	88 (2.8)	0.004
Cardiac			

Outcome

Exposure	+	-	
+	a	b	a + b
-	c	d	c + d

2. Summarize the data

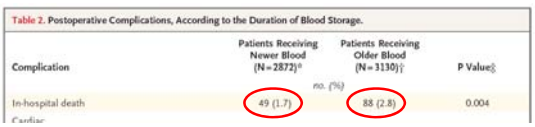

Table 2. Postoperative Complications, According to the Duration of Blood Storage.

Complication	Patients Receiving Newer Blood (N = 2872)*	Patients Receiving Older Blood (N = 3130)†	P Value‡
	no. (%)		
In-hospital death	49 (1.7)	88 (2.8)	0.004
Cardiac			

Death

Old blood	Yes	No	
+	88	3042	3130
-	49	2823	2872

3. Calculate relevant measures of disease occurrence

		Death: Yes	Death: No	
Old blood	+	88	3042	3130
	-	49	2823	2872

- CI_{exp} = 88/3130 = 2.81%
- CI_{unexp} = 49/2872 = 1.74%

4. Calculate measure of association

		Death: Yes	Death: No	
Old blood	+	88	3042	3130
	-	49	2823	2872

- CIR = Ci_{exp}/Ci_{unexp} = (88/3130)/(49/2872) =
- CIR = 1.65

5. Interpret results

- CIR = 1.65
- So?
- Alt 1. The relative risk is 1.65
- Alt 2. The cumulative incidence ratio is 1.65
- Alt 3. The risk of death is 65% higher among recipients of old blood compared to recipients of “young” blood

Example 2 – SES in AML, etc.

JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

Socioeconomic Differences in Patient Survival Are Increasing for Acute Myeloid Leukemia and Multiple Myeloma in Sweden

ABSTRACT

1. Formulate the question

- Does mortality following a diagnosis of AML differ between the SES groups?
- What is the exposure?
  - Socioeconomic status
- What is the outcome?
  - Death within 10 years of AML diagnosis

2. Summarize the data

Socioeconomic status	Deaths	Person years
Blue-collar worker	2301	5918.6
Self-employed	452	880.7
Higher white-collar worker	331	1242.8
Farmer	551	753.6
Lower white-collar worker	1903	4787.3
Retired	316	276.9

What is the relevant measure of disease occurrence?

3. Calculate relevant measures of disease occurrence

Socioeconomic status	Deaths	Person years	Incidence rate/1,000
Blue-collar worker	2301	5918.6	388.8
Self-employed	452	880.7	513.2
Higher white-collar worker	331	1242.8	266.3
Farmer	551	753.6	731.2
Lower white-collar worker	1903	4787.3	397.5
Retired	316	276.9	1141.2

Before calculating, what do we need to decide?

A reference category!

4. Calculate measure of association

Socioeconomic status	Incidence rate/1,000	Incidence rate ratio
Blue-collar worker	388.8	1.00
Self-employed	513.2	1.32
Higher white-collar worker	266.3	0.69
Farmer	731.2	1.88
Lower white-collar worker	397.5	1.02
Retired	1141.2	2.94

Reference

5. Interpret results

- $IRR_{Retired} = 2.94$
- i.e. The relative risk of death is almost 3-fold among retired compared to blue-collar workers?
- Why?