Faculty of Health Sciences

Day 5: binary responses and 2×2 tables

Paul Blanche

Section of Biostatistics, University of Copenhagen

November 11, 2020

Outline

Binary outcome

$$Y = \begin{cases} 1 & \text{event / positive / disease} \\ 0 & \text{no event / negative / non-disease} \end{cases}$$

Binary outcome

$$Y = \begin{cases} 1 & \text{event / positive / disease} \\ 0 & \text{no event / negative / non-disease} \end{cases}$$

Parameters

- Prevalence: proportion of the population with event at fixed time point.

  How many have the disease right now?

- Incidence/hazard rate: number of event relative to time unit:

  How many per year newly acquire the disease?

- Risk: probability that event occurs in given time period:

  How likely will a subject acquire the disease within 1-year?

Statistical inference

Estimating risks and prevalence

$$\widehat{p} = \text{Relative frequency} = \frac{\text{Number of events}}{\text{Number of subjects}} = \frac{x}{n}$$

Confidence limits: normal approximation ("large" n[1])

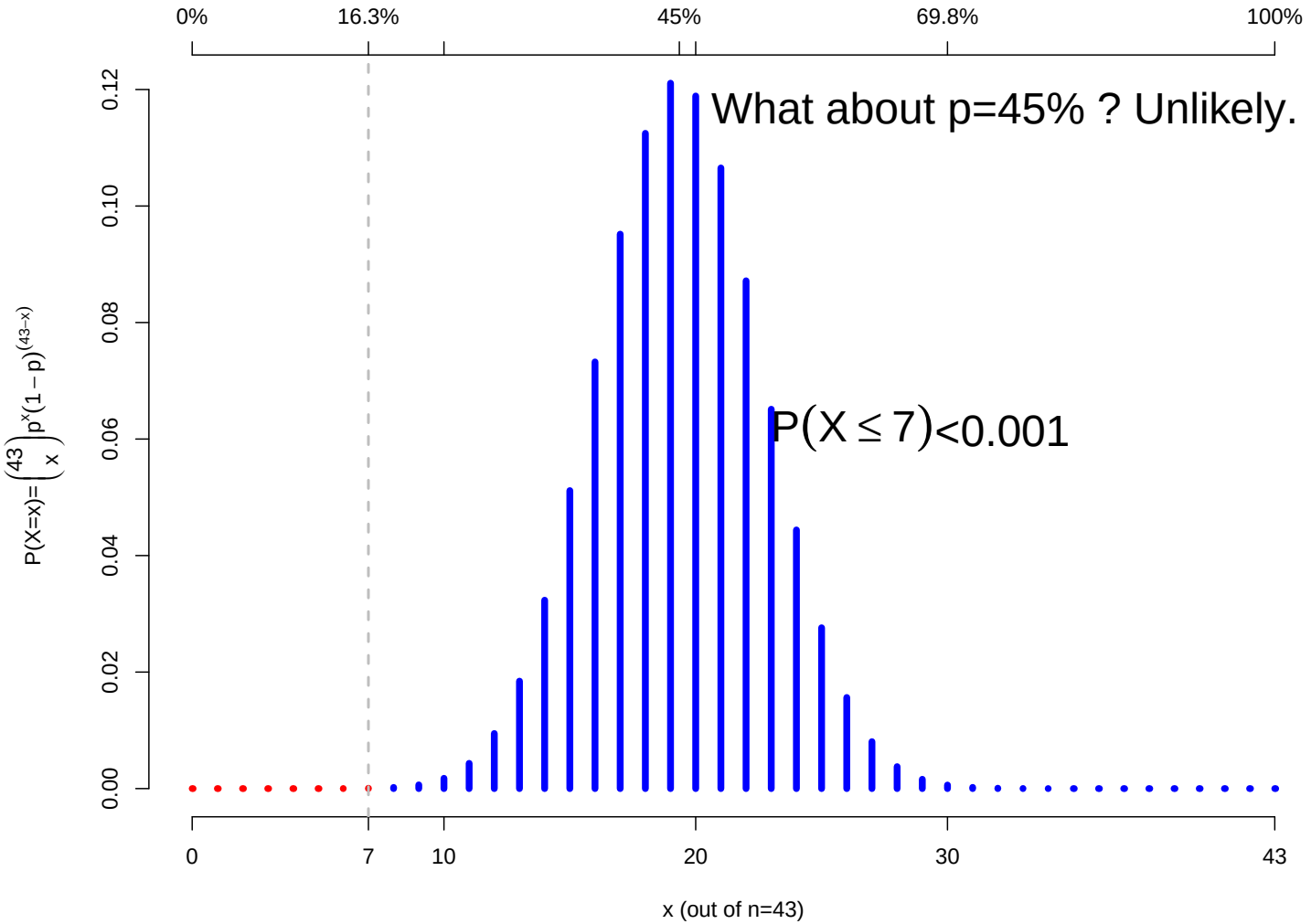
$$\left[\widehat{p} - 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}; \widehat{p} + 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

Confidence limits: "exact" (any n)

```
binom.test(x,n)
```

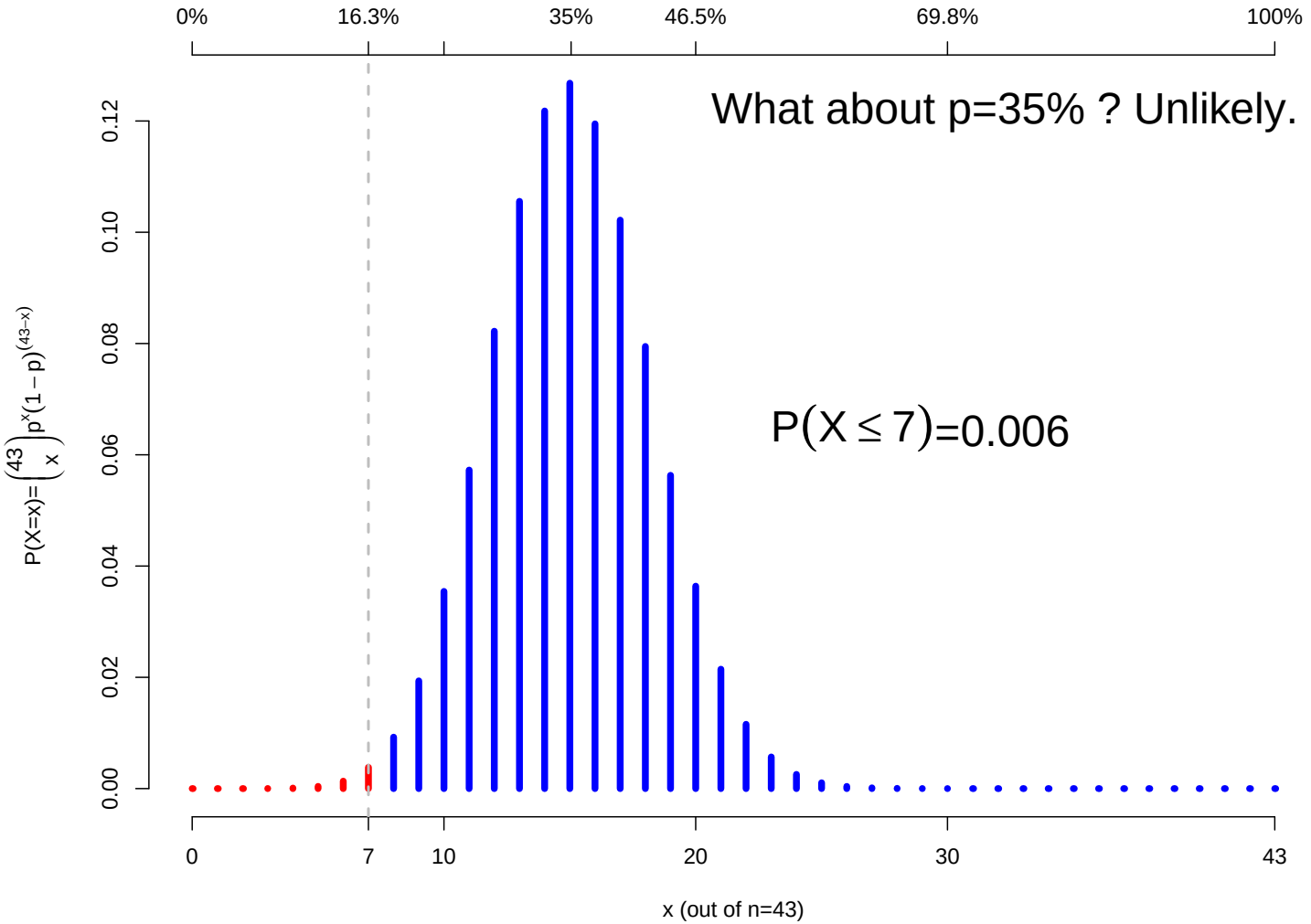
[1] rule of thumb: when both $x \geq 5$ and $n - x \geq 5$.

Exact confidence intervals (computation/intuition)



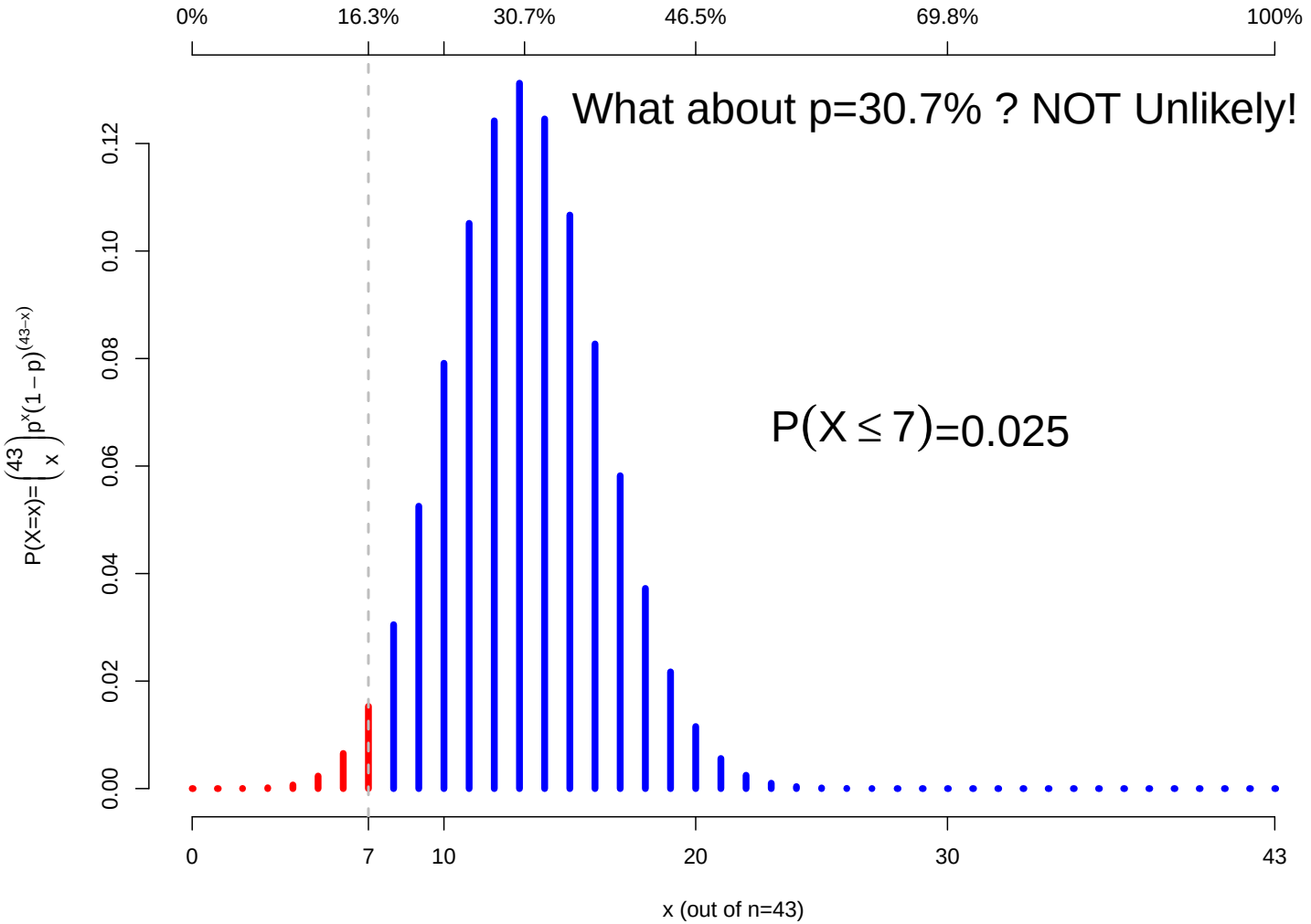
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



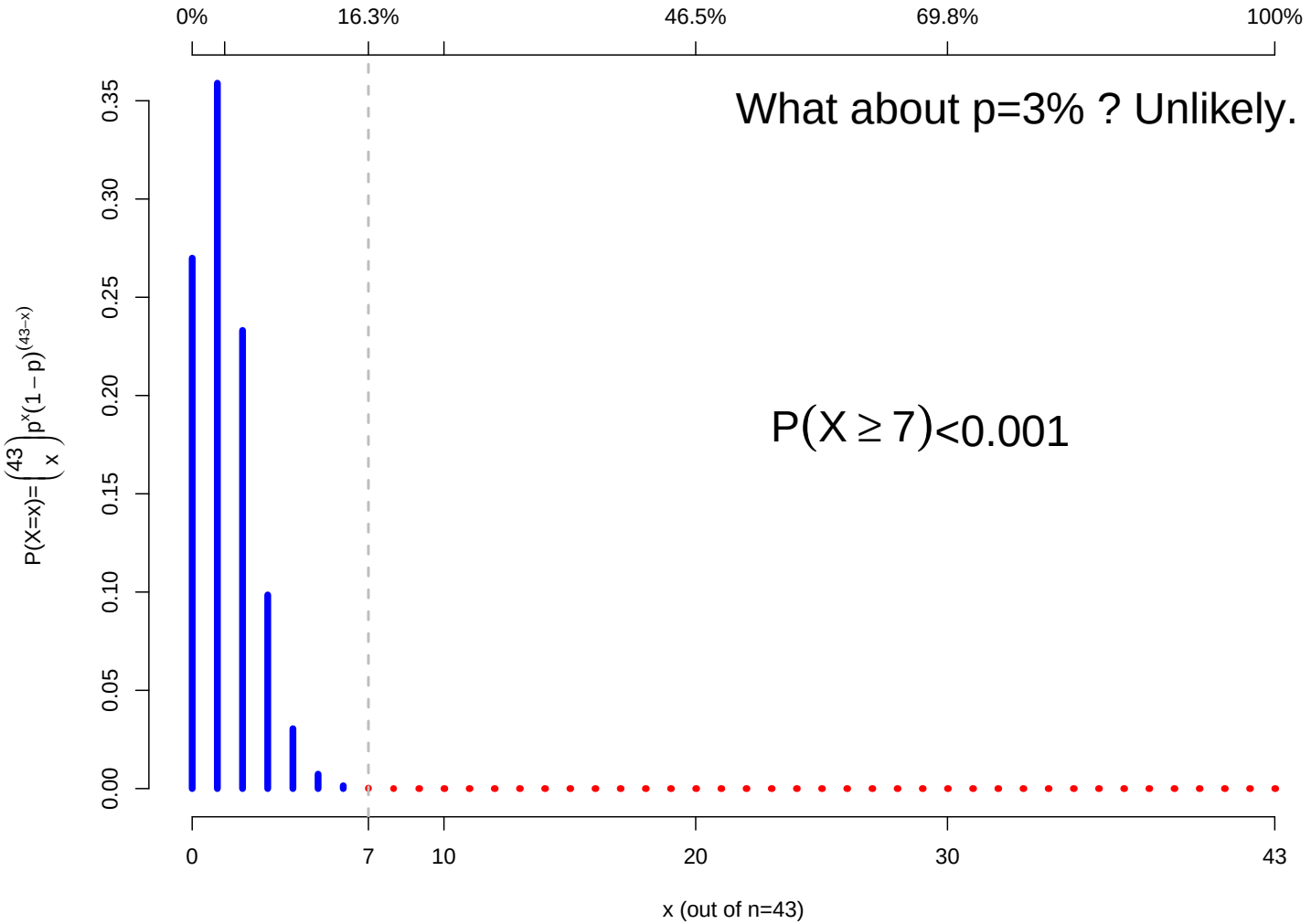
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



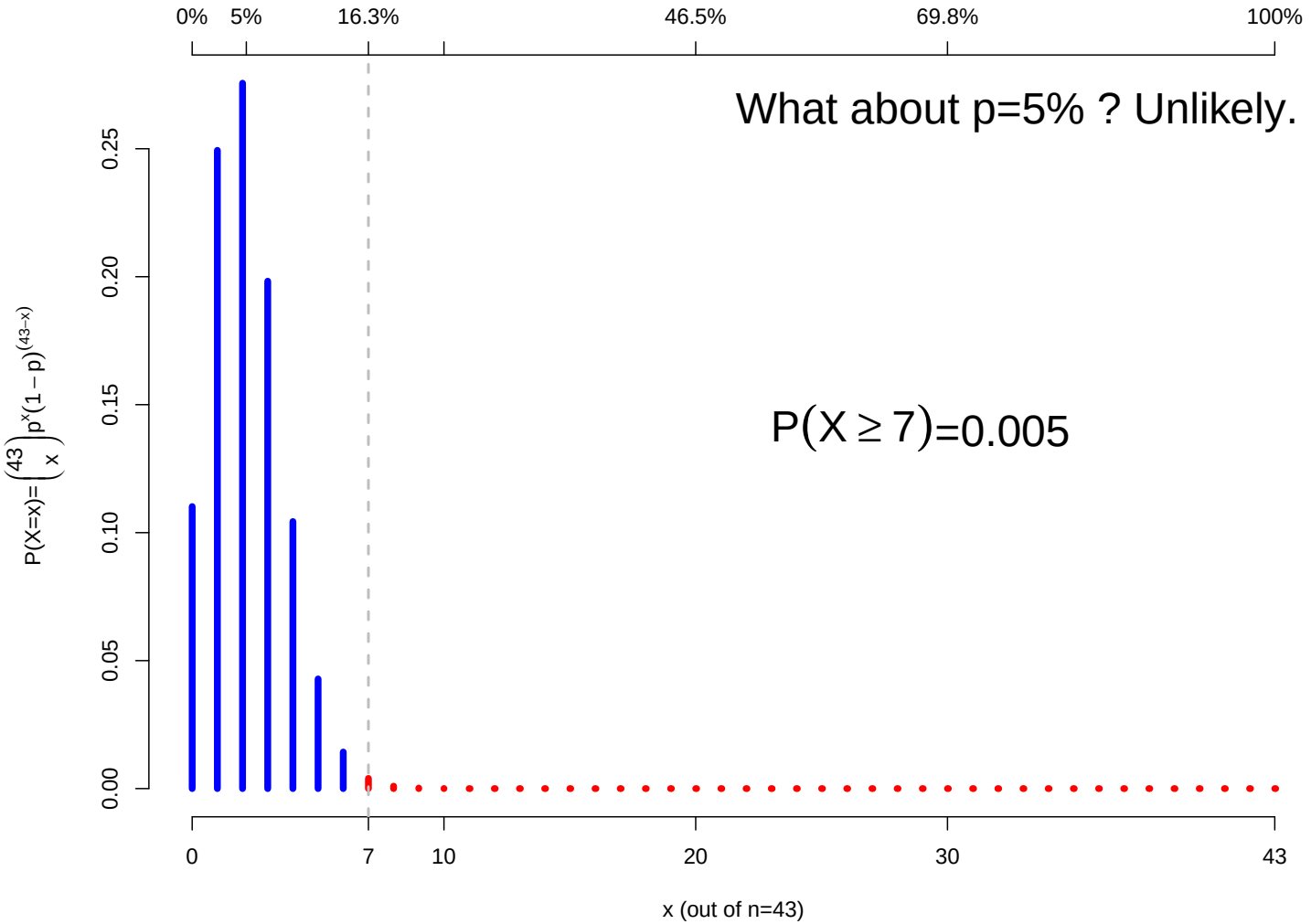
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



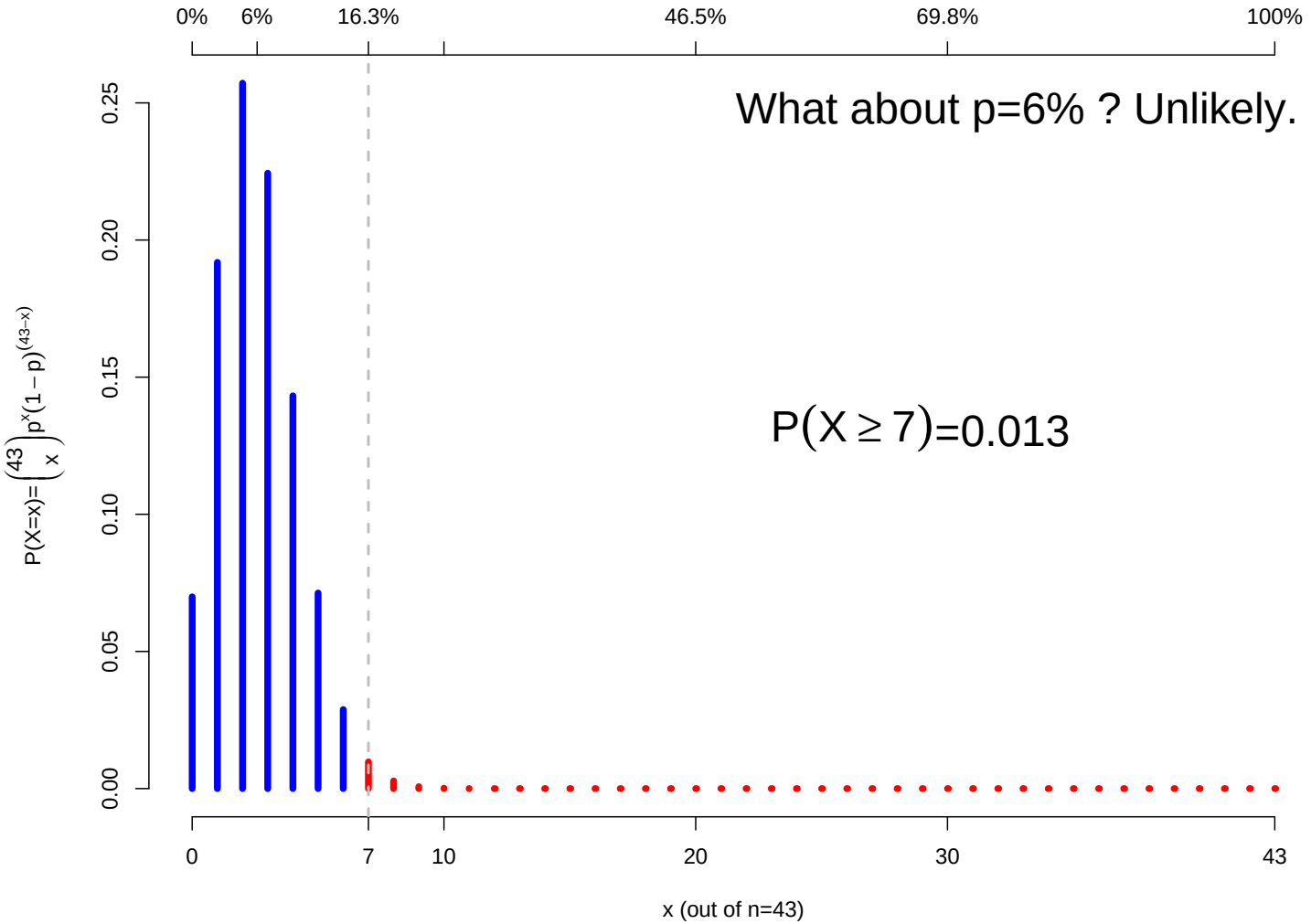
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



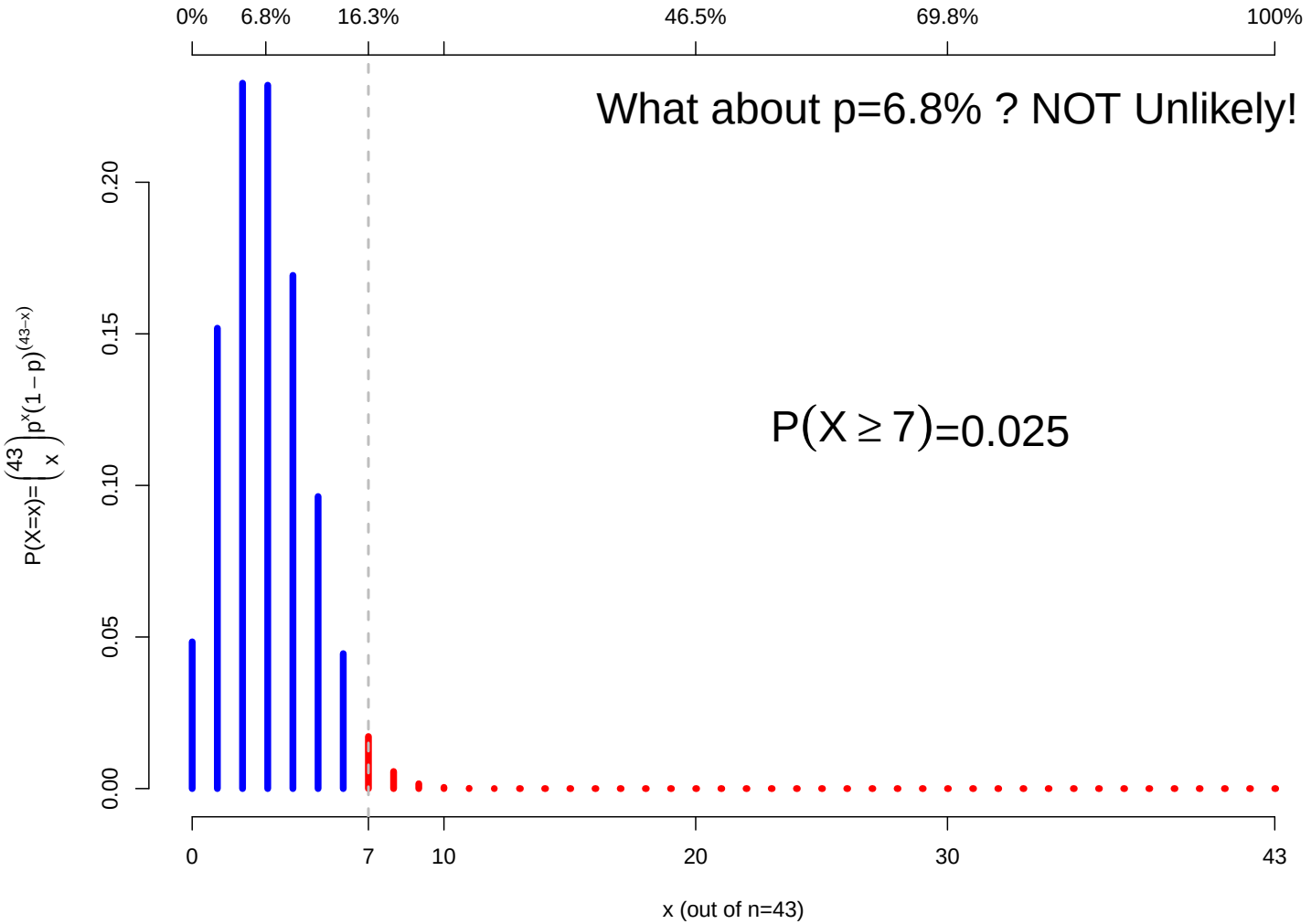
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



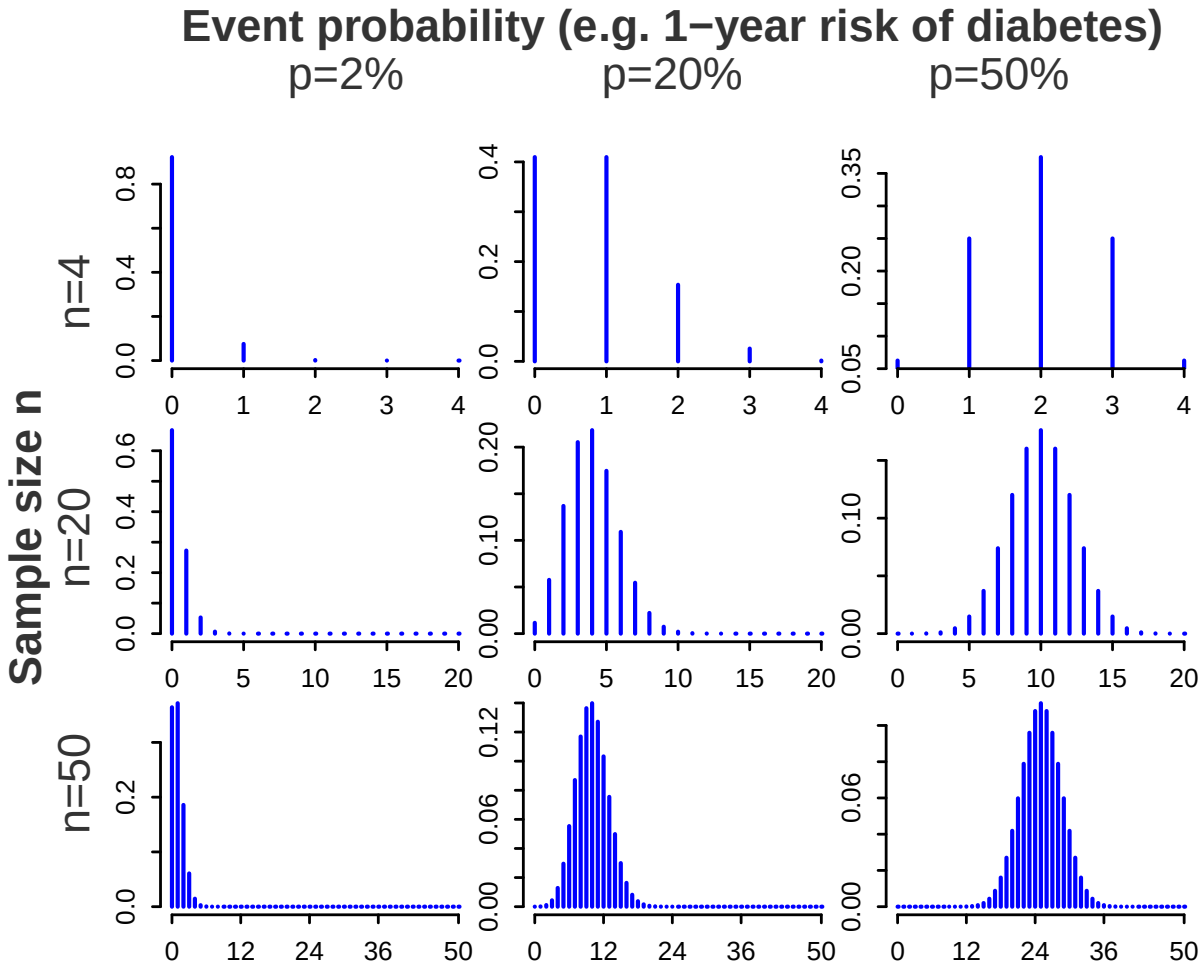
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Exact confidence intervals (computation/intuition)



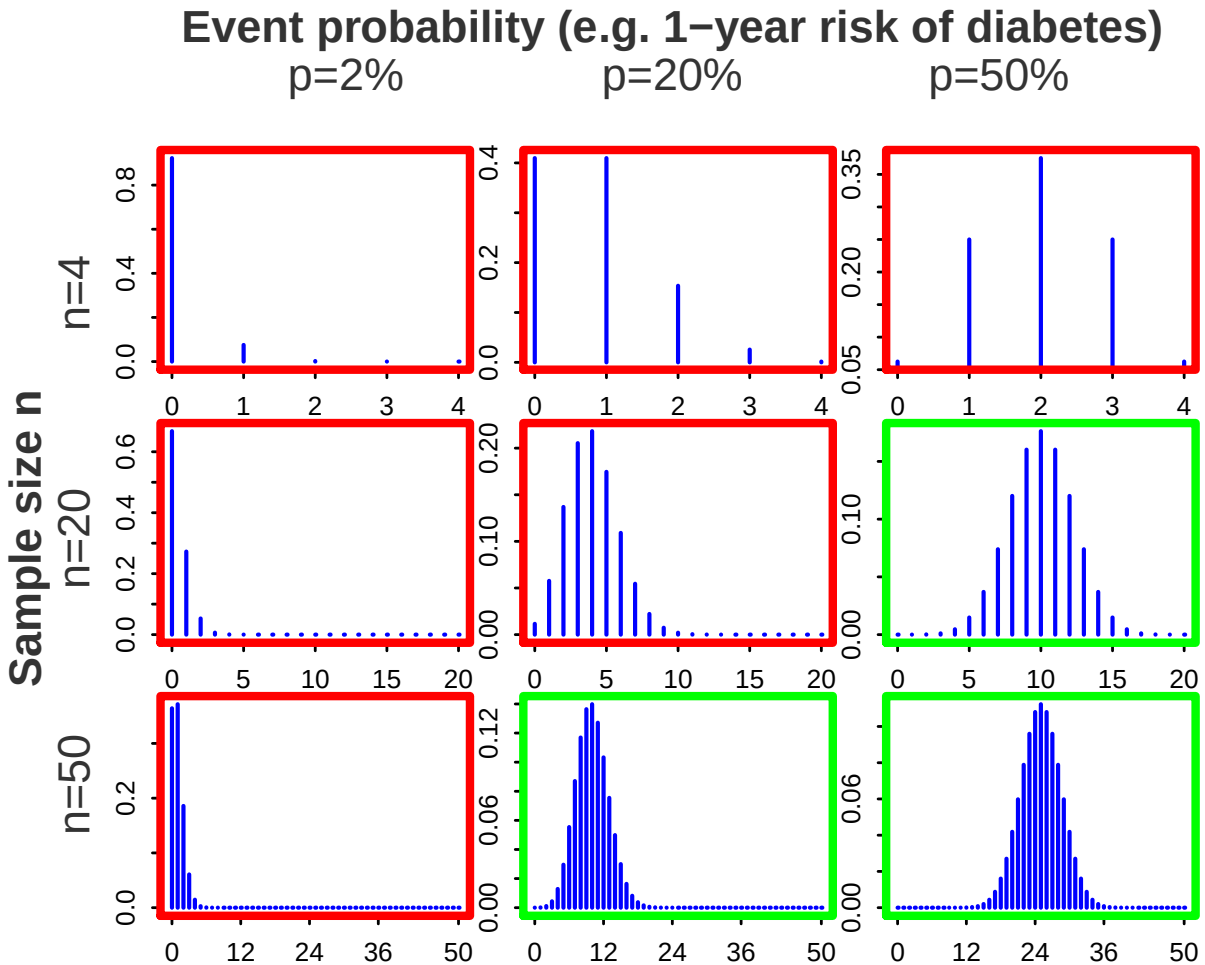
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [6.8; 30.7]$.

Normal approximation



- Binomial distribution: $P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [5.2; 27.3]$.

Normal approximation

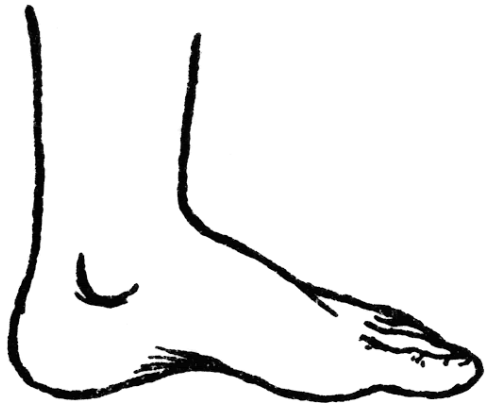


- "good"approximation if $np \geq 5$ and $n(1-p) \geq 5$.
- $x = 7$ and $n = 43$ leads to $\hat{p} = 16.3\%$ and 95% CI$= [5.2; 27.3]$.

Case: clinical trial on Dalteparin [3]

Data: $n = 85$ diabetic patients with peripheral arterial occlusive disease and chronic foot ulcers, randmomized (double-blind) to:

- Placebo ($n = 42$)
- Dalteparin ($n = 43$)



Outcome:

Category [2]	Label
intact skin	healed
decreased ulcer area $\geq$ 50%	improved
increased ulcer area $\geq$ 50%	impaired
decreased or increased ulcer area $<$ 50%	unchanged
amputation above/below ankle	amputation

Research question: Does Dalteparin improve the outcome?

[2] mutually exclusive.

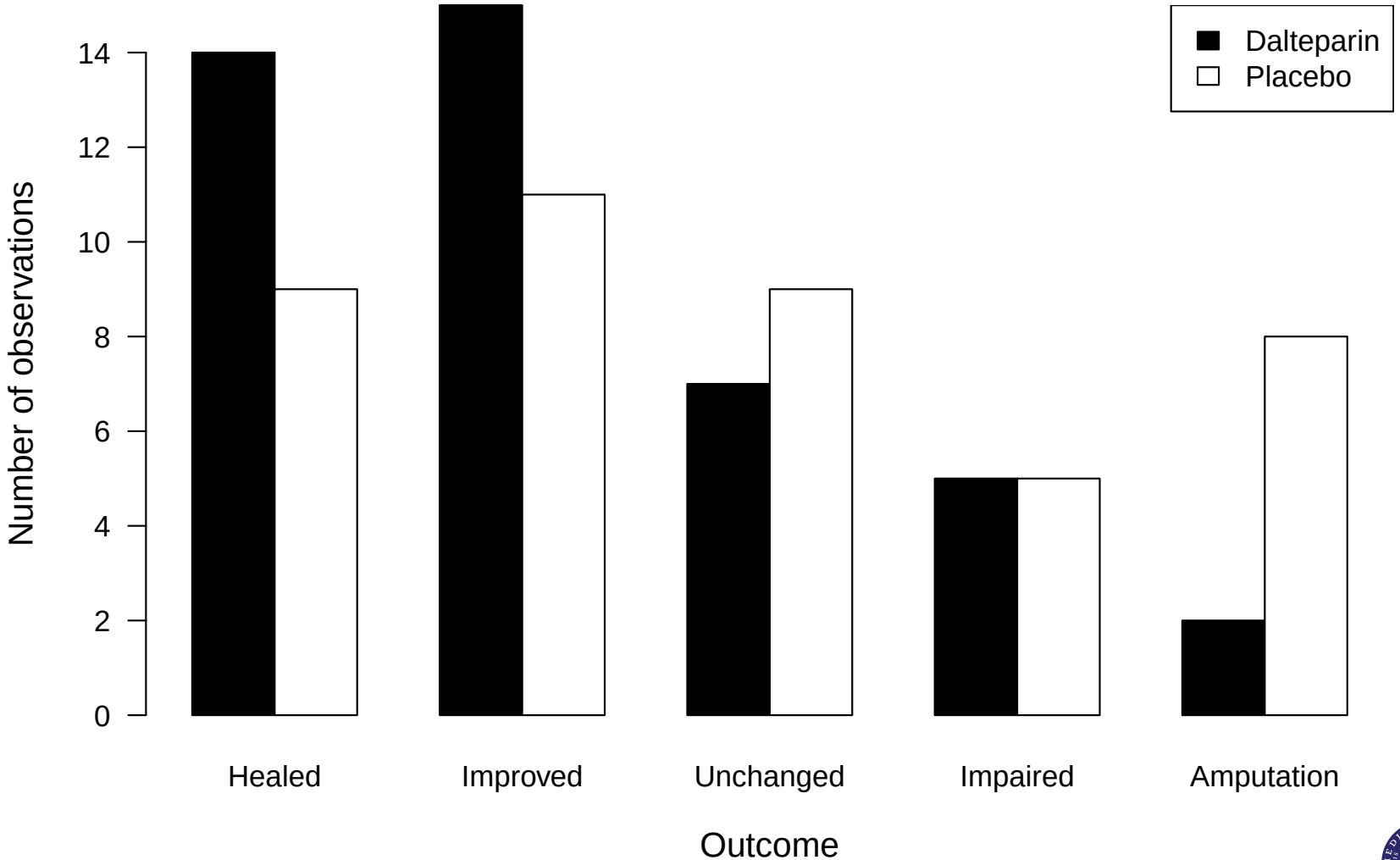
[3] Kalani et al. *Diabetes Care* **26**: 2575-2580, 2003

Frequency table

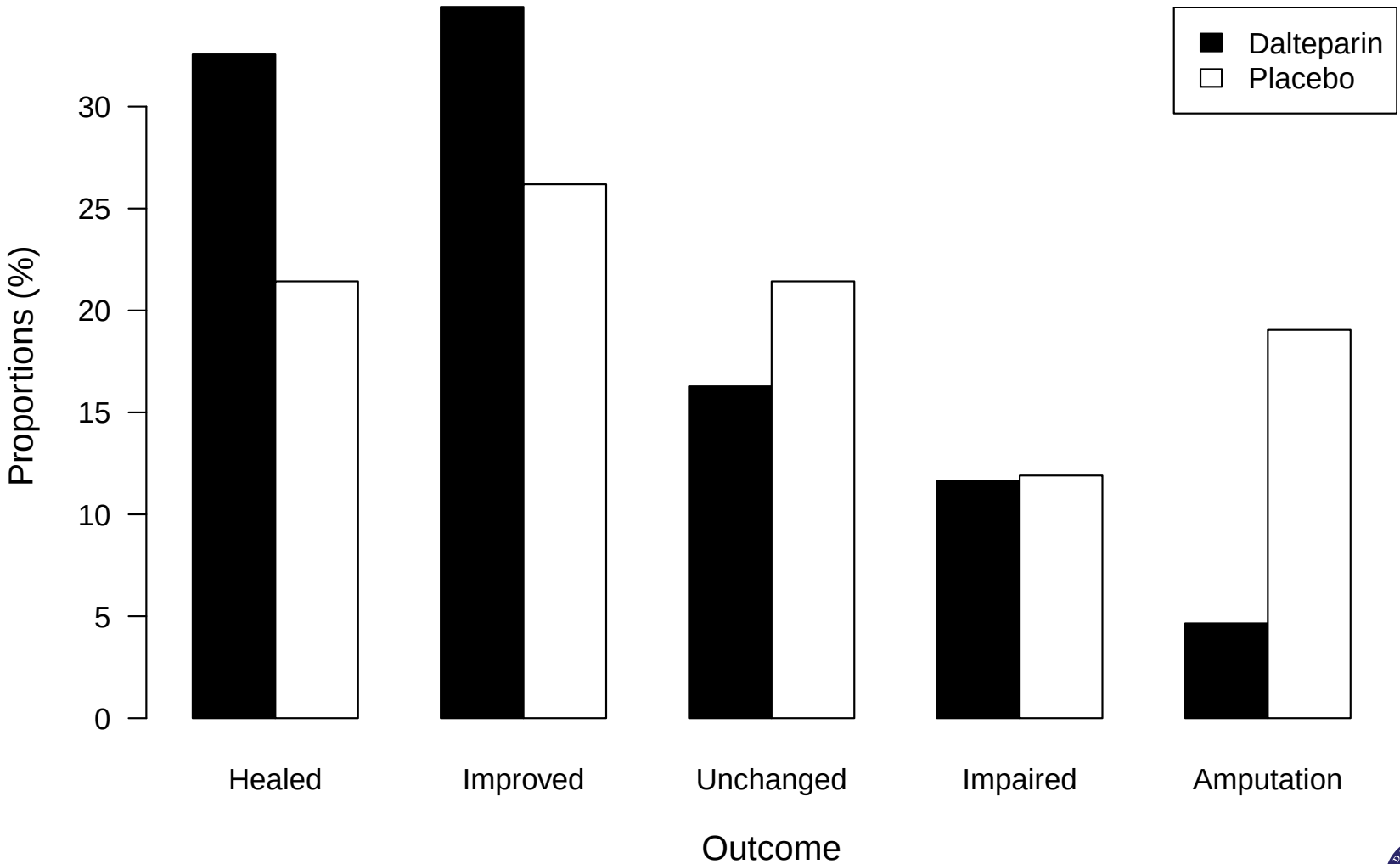
	Dalteparin	Placebo
Healed	14 (33%)	9 (21%)
Improved	15 (35%)	11 (26%)
Unchanged	7 (16%)	9 (21%)
Impaired	5 (12%)	5 (12%)
Amputation	2 (5%)	8 (19%)
total (100%)	43	42

- Summarizes the outcome data.
- Prepare/Format data for analyzes.

Barplot (frequencies)



Barplot (proportions[4])



[4] often better when sample sizes are not equal in both groups.

Here we pool the outcome categories as follows

Category	Dichotomized outcome
intact skin ulcer area decreased $\geq 50\%$	better
decreased or increased ulcer area $< 50\%$ increased ulcer area $\geq 50\%$ amputation above/below ankle	worse

Important: this dichotomization should be prespecified (i.e. decision made before seeing the data).

Outline

Group comparison

Placebo group

$$\text{Risk of worse outcome} = \frac{22}{42} = \widehat{p}_1$$

Dalteparin group

$$\text{Risk of worse outcome} = \frac{14}{43} = \widehat{p}_2$$

[5] whenever possible, we prefer using risk ratios or risk differences to odds ratios. This is so much simpler to understand and to communicate!

Group comparison

Placebo group

$$\text{Risk of worse outcome} = \frac{22}{42} = \widehat{p}_1$$

Dalteparin group

$$\text{Risk of worse outcome} = \frac{14}{43} = \widehat{p}_2$$

Association measures[5]

Relative risk: $\dfrac{\widehat{p}_1}{\widehat{p}_2}$ Odds ratio: $\dfrac{\frac{\widehat{p}_1}{1-\widehat{p}_1}}{\frac{\widehat{p}_2}{1-\widehat{p}_2}}$ Risk difference: $\widehat{p}_1 - \widehat{p}_2$

[5] whenever possible, we prefer using risk ratios or risk differences to odds ratios. This is so much simpler to understand and to communicate!

2x2 contingency table

		Response		
		yes	no	total
Exposure	yes	a	b	a+b
	no	c	d	c+d
	total	a+c	b+d	N

Risk estimates

$$\widehat{p}_1 = \frac{a}{a+b} \qquad \widehat{p}_2 = \frac{c}{c+d}$$

Relative risk

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)}$$

Exposure	Response: yes	Response: no	total
yes	a	b	a+b
no	c	d	c+d
total	a+c	b+d	N

Standard error of $\log(\widehat{RR})$ and confidence interval [6]

$$\widehat{\sigma} = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

$$CI_{95\%} = \left[\widehat{RR} \cdot \exp(-1.96\,\widehat{\sigma}) \;;\; \widehat{RR} \cdot \exp(1.96\,\widehat{\sigma})\right]$$

[6] This method is "good enough" with "large enough" sample sizes.

Relative risk: placebo versus dalteparin

$$\widehat{RR} = \frac{22/42}{14/43} = 1.609$$

Treatment	Outcome: worse	Outcome: better	total
placebo	22	20	42
dalteparin	14	29	43
total	36	49	85

Standard error of $\log(\widehat{RR})$ and confidence interval

$$\hat{\sigma} = \sqrt{\frac{1}{22} - \frac{1}{42} + \frac{1}{14} - \frac{1}{43}} = 0.264$$

$$CI_{95\%} = [0.959; 2.7] \text{ (does include 1)}$$

Risk difference

$$\widehat{\Delta} = \frac{a}{a+b} - \frac{c}{a+b}$$

Exposure	Response: yes	Response: no	total
yes	a	b	a+b
no	c	d	c+d
total	a+c	b+d	N

Standard error of $\widehat{\Delta}$ and confidence interval [7]

$$\widehat{\sigma} = \sqrt{ab/(a+b)^3 + cd/(c+d)^3}$$

$$CI_{95\%} = \left[\widehat{\Delta} - 1.96\,\widehat{\sigma}\ ;\ \widehat{\Delta} - 1.96\,\widehat{\sigma}\right]$$

[7] This method is "good enough" with "large enough" sample sizes.

Risk difference: placebo versus dalteparin

$$\widehat{\Delta} = \frac{22}{42} - \frac{14}{43} = 0.198$$

Treatment	Outcome: worse	Outcome: better	total
placebo	22	20	42
dalteparin	14	29	43
total	36	49	85

Standard error of $\widehat{\Delta}$ and confidence interval

$$\widehat{\sigma} = \sqrt{22 \cdot 20/42^3 + 14 \cdot 29/43^3} = 0.105$$

$$CI_{95\%} = [-0.008 \; ; \; 0.404] \text{ (does include 0)}$$

Odds Ratio (OR)

Concept **needed** for

- case-control studies
- logistic regression

Odds: are defined as "risk of event divided by risk of no event"

$$\boxed{\text{odds} = p/(1-p)},$$

and the risk can be computed back from the odds, $p = \text{odds}/(1+\text{odds})$.

Odds are difficult to interpret, but if risks are small, then risks $\approx$ odds.

The Odds ratio (OR) is defined as the ratio of the odds,

$$\boxed{OR = \frac{\text{odds}_1}{\text{odds}_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}}.$$

OR are difficult to interpret, but from the equation...

$$RR = \frac{OR}{\left\{1-p_2\right\} + p_2 OR},$$

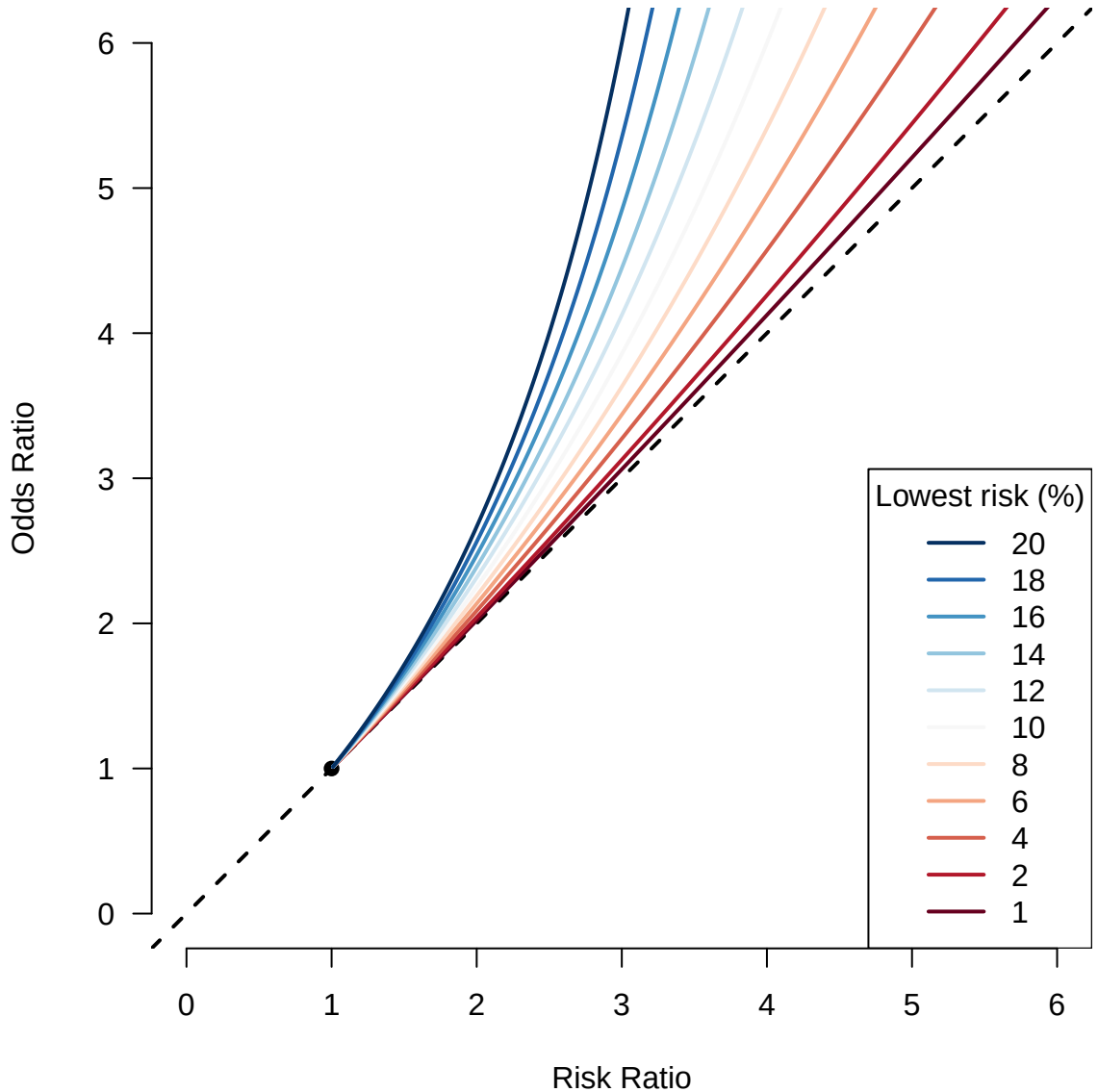
...we can first conclude:

- $OR > 1 \Leftrightarrow RR > 1$
- $OR = 1 \Leftrightarrow RR = 1$
- $OR < 1 \Leftrightarrow RR < 1$

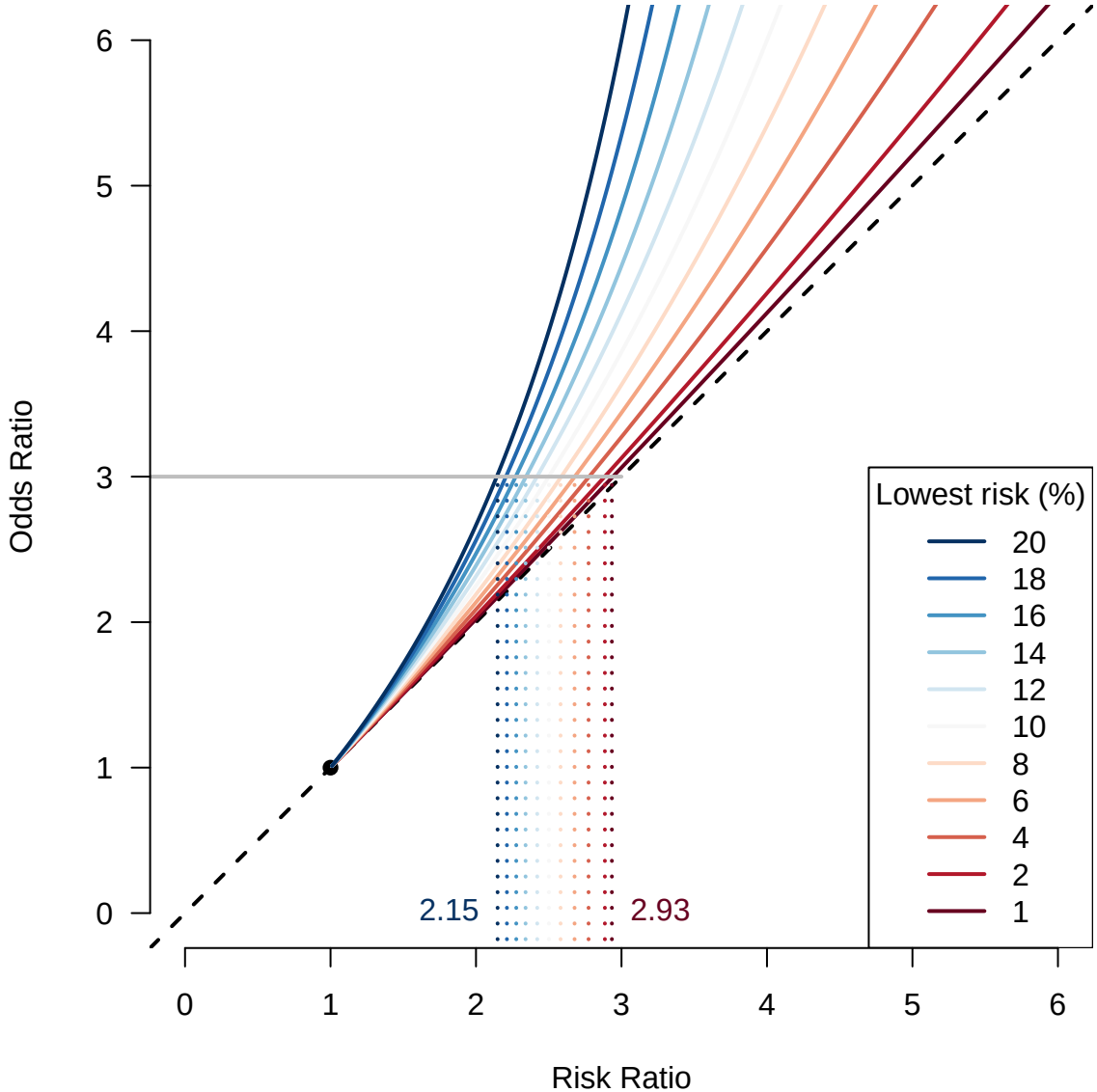
...and further conclude that

- the OR is sufficient to deduce whether a risk increases or decreases.
- if p_2 is small (e.g. rare disease), then $OR \approx RR$.

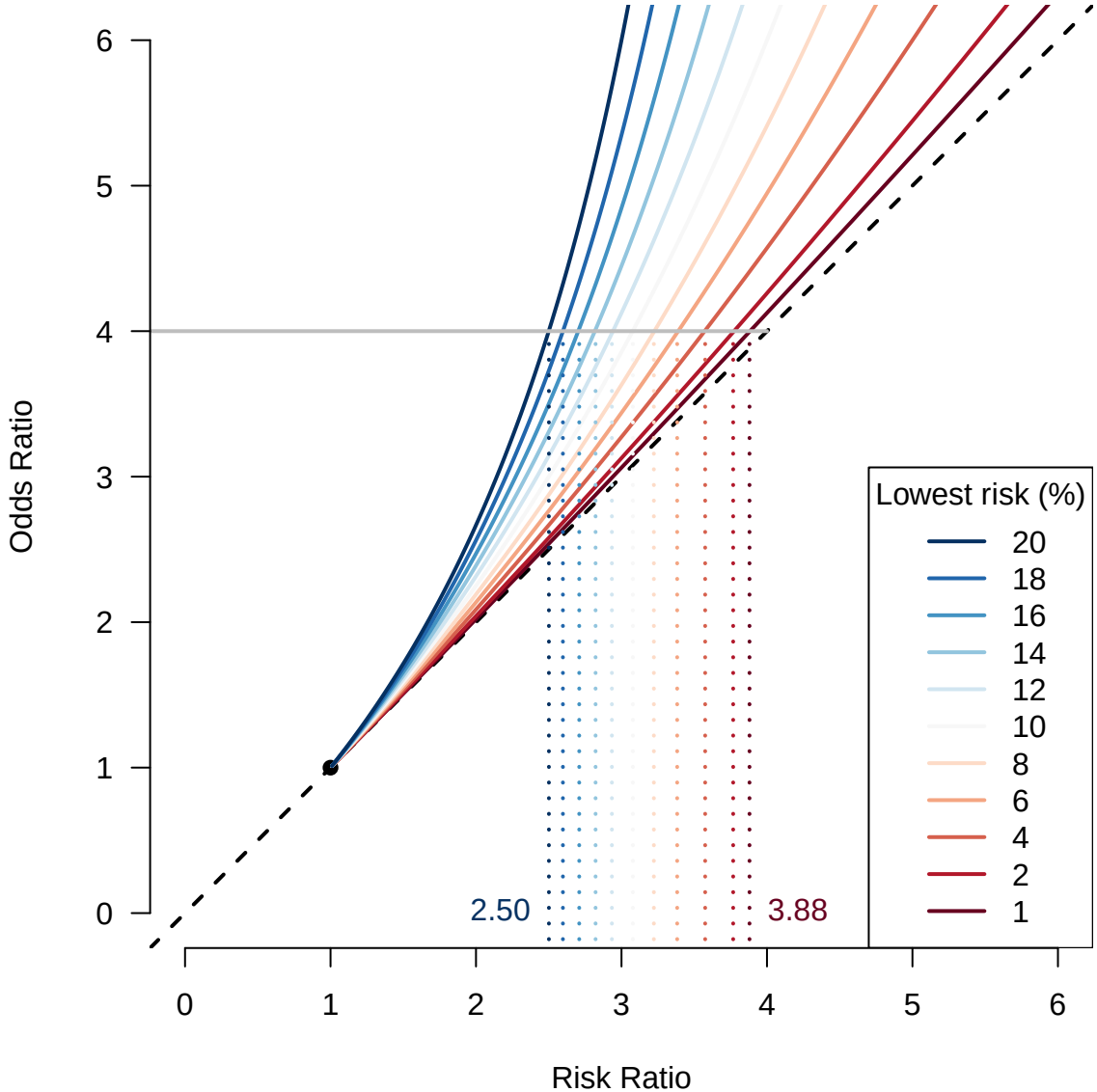
When is $OR \approx RR$?



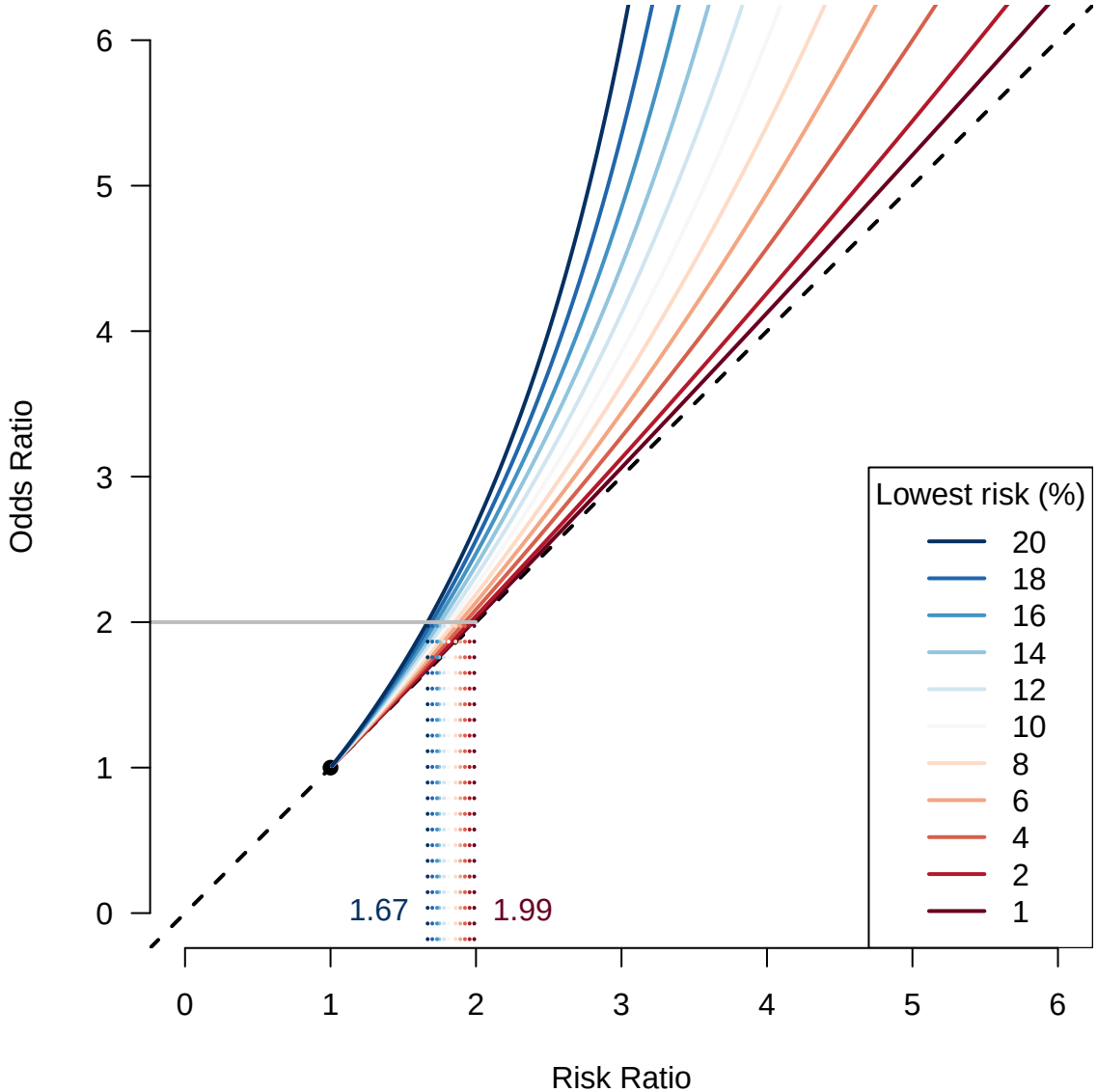
When is $OR \approx RR$?



When is $OR \approx RR$?



When is $OR \approx RR$?



Odds ratio

$$\widehat{OR} = \frac{\frac{a/(a+b)}{b/(a+b)}}{\frac{c/(c+d)}{d/(c+d)}} = \frac{a \cdot d}{b \cdot c}$$

Exposure \ Response	yes	no	total
yes	a	b	a+b
no	c	d	c+d
total	a+c	b+d	N

Standard error of $\log(\widehat{OR})$ and confidence interval[8]

$$\widehat{\sigma} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$CI_{95\%} = \left[\widehat{OR} \cdot \exp(-1.96\,\widehat{\sigma}); \widehat{OR} \cdot \exp(1.96\,\widehat{\sigma})\right]$$

[8] This method is "good enough" with "large enough" sample sizes.

Odds ratio: placebo versus dalteparin

$$\widehat{OR} = \frac{22 \cdot 29}{14 \cdot 20} = 2.279$$

Treatment	Outcome: worse	Outcome: better	total
placebo	22	20	42
dalteparin	14	29	43
total	36	49	85

Standard error of $\log(\widehat{OR})$ and confidence interval

$$\widehat{\sigma} = \sqrt{\frac{1}{22} + \frac{1}{20} + \frac{1}{14} + \frac{1}{29}} = 0.449$$

$$CI_{95\%} = [0.946; 5.491]$$ (does include 1)

Reporting results

The relative risk (of worsening) of group 1 (Dalteparin) versus group 2 (Placebo) is estimated as

$$RR = \frac{14/43}{22/42} = 0.622$$

Equivalent statements:

- *The risk in group 2 is reduced by a factor 0.622 compared to group 1.*
- *The risk in group 2 is* **37.8% lower** *than in group 1.*[9]

- *The risk in group 1 is 1.609 times higher than in group 2.*[10]
- *The risk in group 1 is* **60.9% higher** *than in group 2.*

[9] because 1-0.622=0.378

[10] because 1/0.622=1.609

Testing independence in a randomized clinical trial

Null hypothesis: the treatment has no effect.

$$
\begin{aligned}
&\text{Prob(worse given dalteparin)} = \text{Prob(worse given placebo)} \\
\Leftrightarrow \quad & p_1 - p_2 = 0 && \text{(Difference =0)} \\
\Leftrightarrow \quad & \frac{p_1}{p_2} = 1 && \text{(Relative risk =1)} \\
\Leftrightarrow \quad & \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = 1 && \text{(Odds ratio =1)}
\end{aligned}
$$

Popular tests of independence between the treatment group and the outcome groups:

- χ^2 test (normal approximation)[11]
- Fisher's exact test

[11] This method is "good enough" with "large enough" sample sizes.

The χ^2 test statistic

$$\chi^2 = \sum \frac{(\text{observed counts} - \text{expected counts})^2}{\text{expected counts}}$$

Observed counts

		Response		
		yes	no	total
Exposure	yes	a	b	a+b
	no	c	d	c+d
	total	a+c	b+d	N

Expected counts

		Response		
		yes	no	total
Exposure	yes	(a+b)(a+c)/N	(a+b)(b+d)/N	a+b
	no	(c+d)(a+c)/N	(c+d)(b+d)/N	c+d
	total	a+c	b+d	N

The expected counts are calculated under the null hypothesis.

Rule of thumb: a valid analysis requires that all expected counts are ≥ 5.

- under the null hypothesis the groups are identical, hence data can be merged into a single group
- in a population of size n, for a given risk of event p, we expect to see (on average) np events in this population

Test results

Null hypothesis:
dalteparin treatment has no effect for chronic foot ulcers.

Test	p-value
Fisher's exact test	0.0808
Pearson's χ^2 test	0.0644
Pearson's χ^2 test with Yates' continuity correction[12]	0.1032

R code:

```
tab <- rbind(c(22,20),c(14,29))
fisher.test(tab)
chisq.test(tab,correct=FALSE)
chisq.test(tab,correct=TRUE)
```

[12] Expected to be more precise than the usual Pearson's χ^2 test when the sample size is very small.

A note of caution

Because the (simple) formulas provided above for the 95% CI are based on large sample size approximations, they are not necessarily consistent with the result of the Fisher's exact test, with "very small" sample sizes.

Example:

	event	no event
exposed	5	12
non-exposed	8	3

- $\widehat{p}_1 = 8/11 = 0.73, \quad \widehat{p}_2 = 5/17 = 0.29.$
- $\widehat{\Delta}$= 0.43 (0.09 ; 0.77)
- $\widehat{RR}$= 2.47 (1.09 ; 5.62)
- $\widehat{OR}$= 6.40 (1.18 ; 34.61)
- p-values from Fisher's exact test and Pearson's χ^2 (with and without Yates correction) are 0.051, 0.063 and 0.025, respectively.

Here the confidence intervals show a significant result, but not Fisher's test.

Advanced methods and software[13] are available to avoid running into this kind of inconsistency between hypothesis test and confidence intervals.

Fortunately, it is rare that we run into this problem....
and even rarer that it matters for the interpretation.

[13] see R package `exact2x2` and references in the help documentation.

Larger contigency tables (1/2)

If the table is not 2x2 but, e.g., 3x4 or 2x4, the χ^2 test and Fisher's exact test are testing an "ANOVA-like" null hypothesis similarly to what the F-test does to compare several means.

First example:

	underweight	normal	overweight	obese
no SCD	9	51	20	8
SCD	23	61	3	1

R code:

```
fisher.test(table(d$SCD,d$BMIgroup))
```

returns a p-value <0.001, for the null hypothesis

H_0: "the prevalence of SCD is the same in all BMI groups"

(i.e. "no association between BMI and SCD").

Larger contigency tables (2/2)

Second example:

	underweight	normal	overweight	obese
age=$[16, 25)$	14	45	1	1
$[25, 30)$	3	25	3	1
$[30, 67]$	15	42	19	7

R code:

```
fisher.test(table(d$ageGroup,d$BMIgroup))
```

returns p-value=0.004, for the null hypothesis

H_0: "the prevalence of each BMI group is the same in all age groups"

(i.e. "no association between BMI and age").

Outline

Textbook formula ("large n" approximation)

$$n = \frac{\left\{ z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + z_{\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)} \right\}^2}{(p_1-p_2)^2}$$

- $z_{\alpha/2} = -1.96$ for $\alpha = 5\%$.[14]
- $z_{1-\beta} = 0.84$ and 1.28 for $1-\beta = 80\%$ and 90%.
- $\bar{p} = (p_1+p_2)/2$.
- n: number of observations in **each** group.

Useful for computing:

- Sample size: n for given "guesses" of p_1 and p_2 and desired $1-\beta$ and α.
- Power for a given budget/sample size: $1-\beta$ for "guesses" of p_1 and p_2 and desired n and α.
- Least detectable difference (or ratio): $\delta = p_1 - p_2$ (or $r = p_1/p_2$) for given n, "guess" of p_1 and desired α and minimal power $1-\beta$.

[14] where z_γ is the γ-quantile of a standard normal distribution.

Sample size calculation

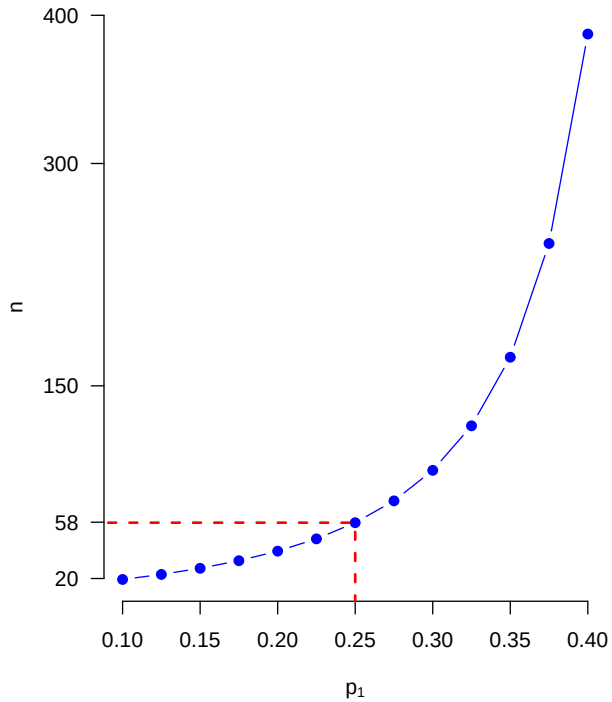
Standard software can be used, e.g. R:

```
power.prop.test(p1 = 0.25, p2 = 0.5, power=0.8)

     Two-sample comparison of proportions power calculation

              n = 57.67344
             p1 = 0.25
             p2 = 0.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```



- $n = 58$ subjects needed in **each** group (i.e. 116 in total) to detect significant risk difference with a power of 80%, if the risks in the two groups are 25% and 50%.

Power calculation

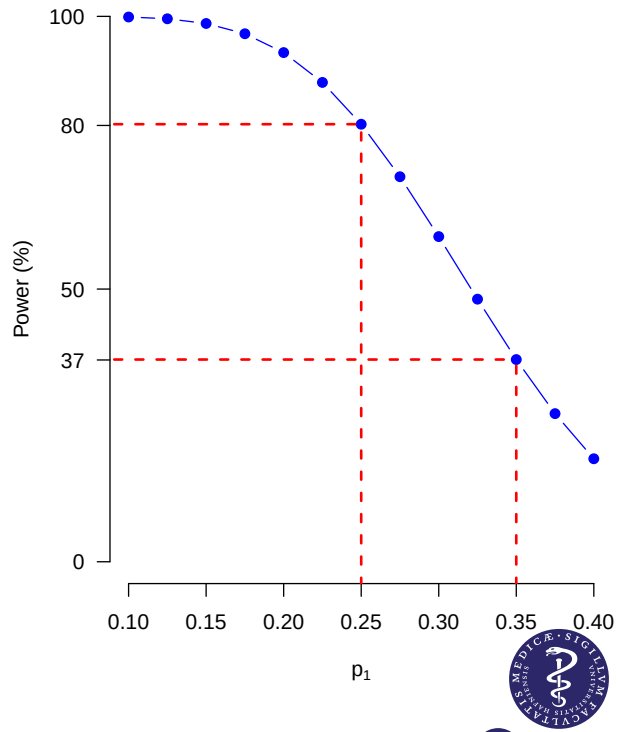
Example: an initial calculation suggests $n = 58$ subjects per group (i.e. 116 in total), for detecting a difference of 25% survival between the two groups, assuming 50% survival in the placebo group (with 80% power). But what does the power become if we were too optimistic with the expected treatment effect? E.g. what if the difference in survival probability is only 15%?

```
power.prop.test(n=58, p1 = 0.35, p2 = 0.5)

     Two-sample comparison of proportions power calculation

              n = 58
             p1 = 0.35
             p2 = 0.5
      sig.level = 0.05
          power = 0.3707966
    alternative = two.sided

NOTE: n is number in *each* group
```



Least detectable difference

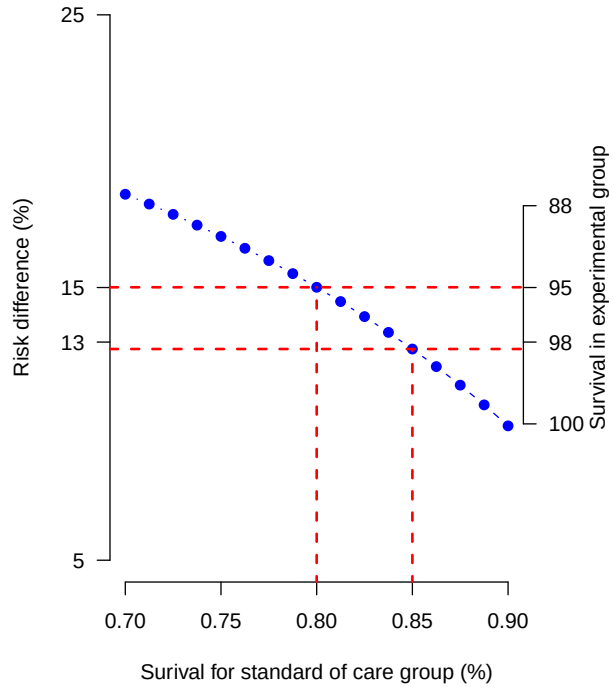
Example: My grant can finance a total sample size of $n = 150$ (i.e. 75 per group). What is the smallest survival difference that I can hope to show with a decent power (e.g. 80%), if I expect 80% survival in the "standard of care" (i.e. control) group? And if I expect 85% in the "standard of care" group?

```
power.prop.test(n=75, p1 = 0.8, power=0.8)

     Two-sample comparison of proportions power calculation

              n = 75
             p1 = 0.8
             p2 = 0.950095
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```



Note: you need to supply a value for `p1`, not `p2`, otherwise the software is looking for a lower risk and it returns 0.72.

Outline

Confounding

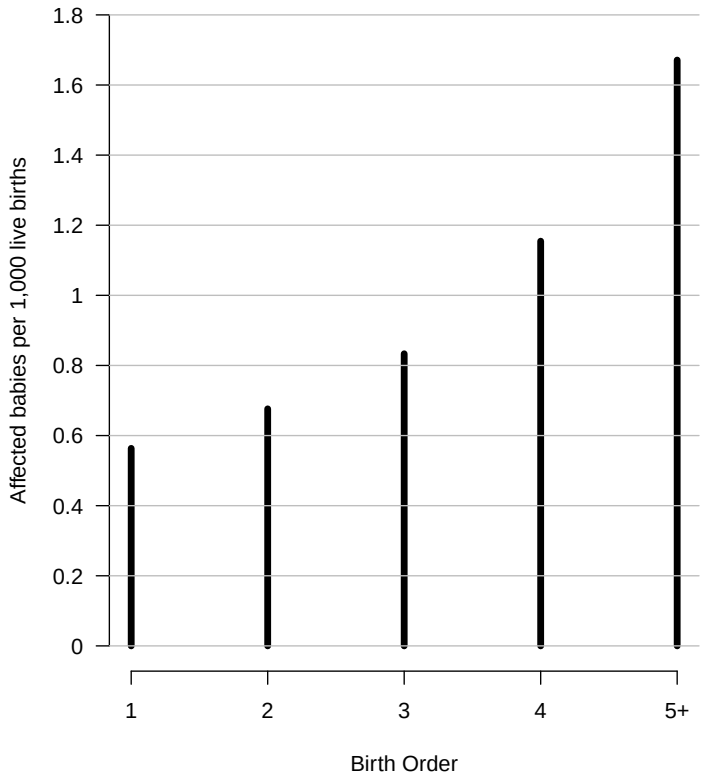
"A simple definition of confounding is the confusion of effects. This definition implies that the effect of the exposure is mixed with the effect of another variable, leading to a bias."[15]

Failing to take a confounding variable into account can lead to a **false conclusion** that the outcome are in a **causal** relationship with the predictor variable.

Confounding variables are typically encountered in observational studies, but **not** in "ideal" randomized experiments.

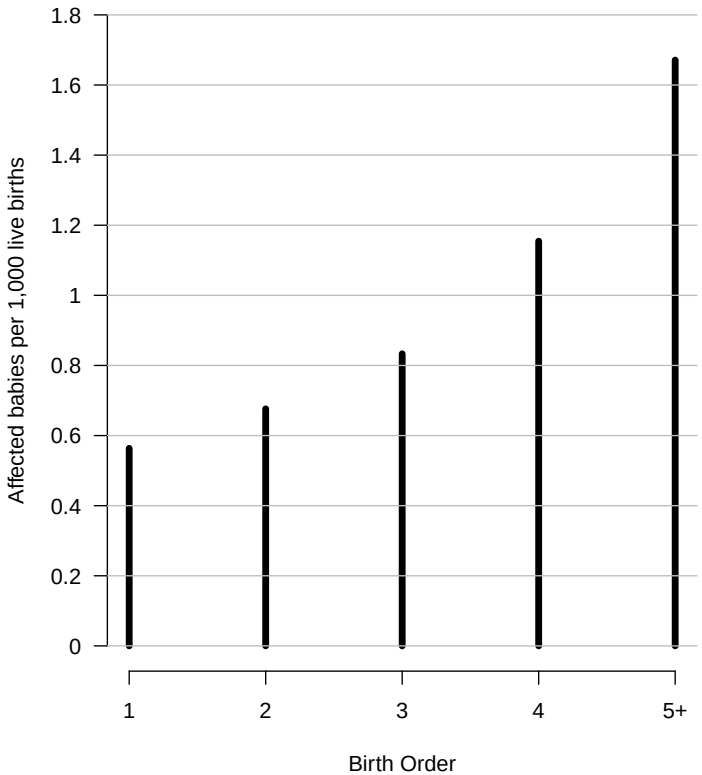
[15] Rothman (2012), Epidemiology: an introduction.

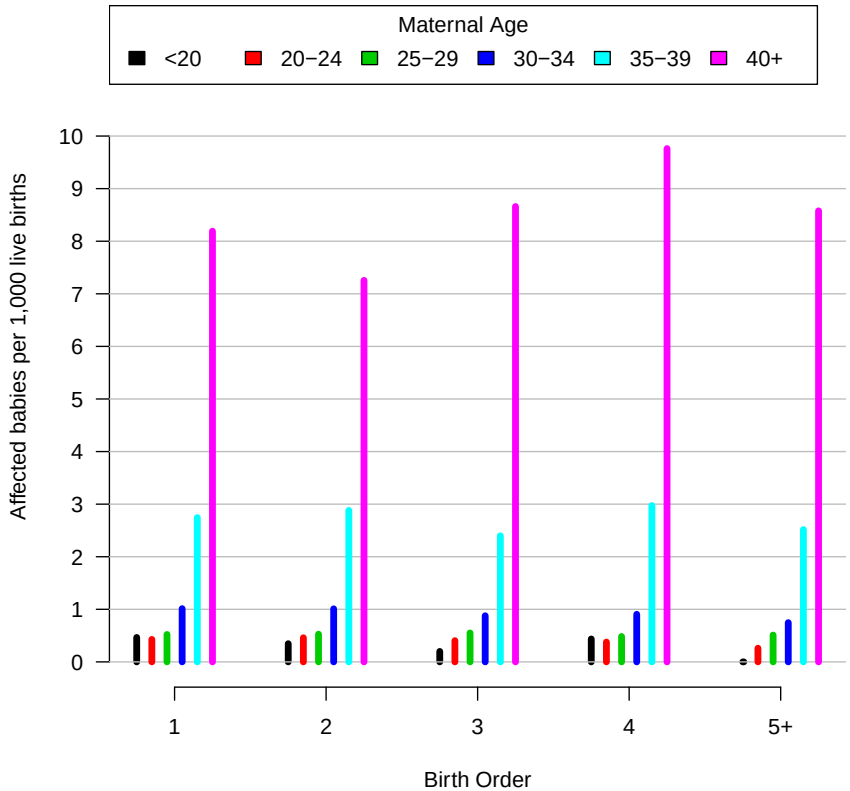
Confounding example (birth order and risk of mongolism [16])



[16] Stark and Mantel (1966) 'Effects of maternal age and birth order on the risk of mongolism and leukemia' J Natl Cancer Inst37(5) 687–698.

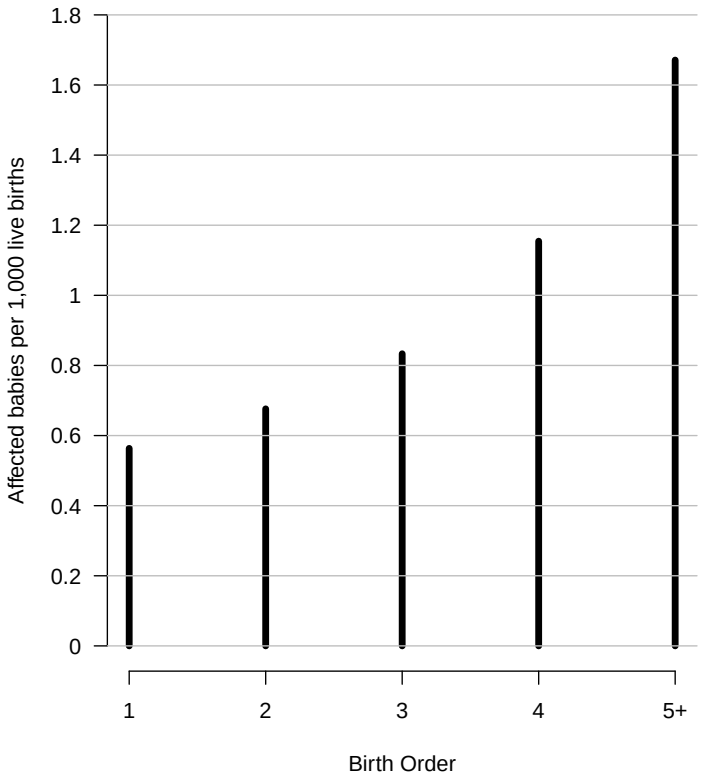
Confounding example (birth order and risk of mongolism [16])

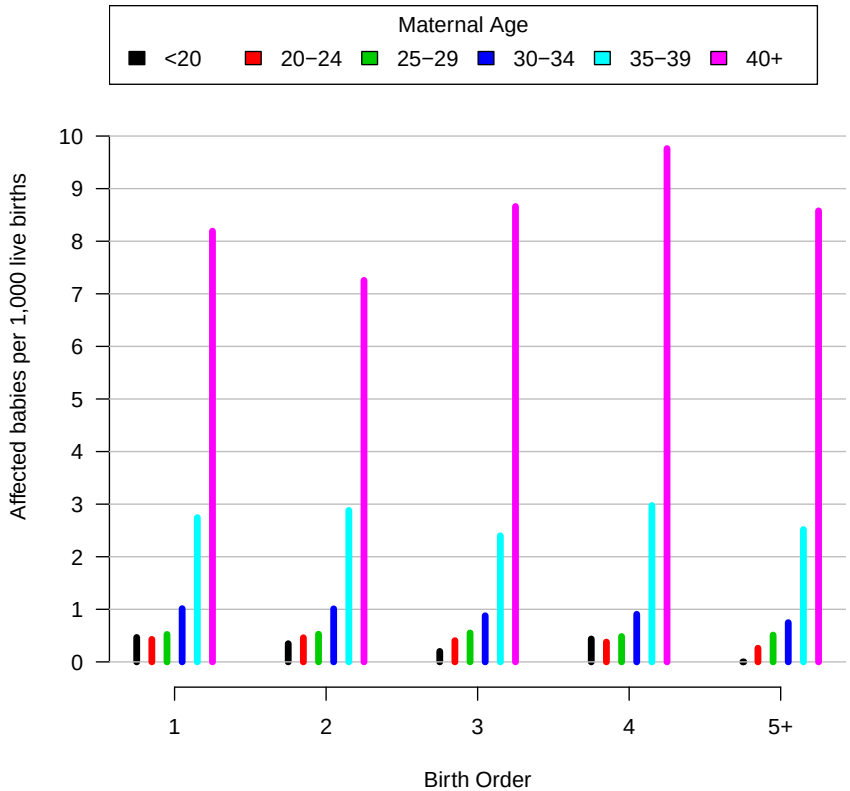




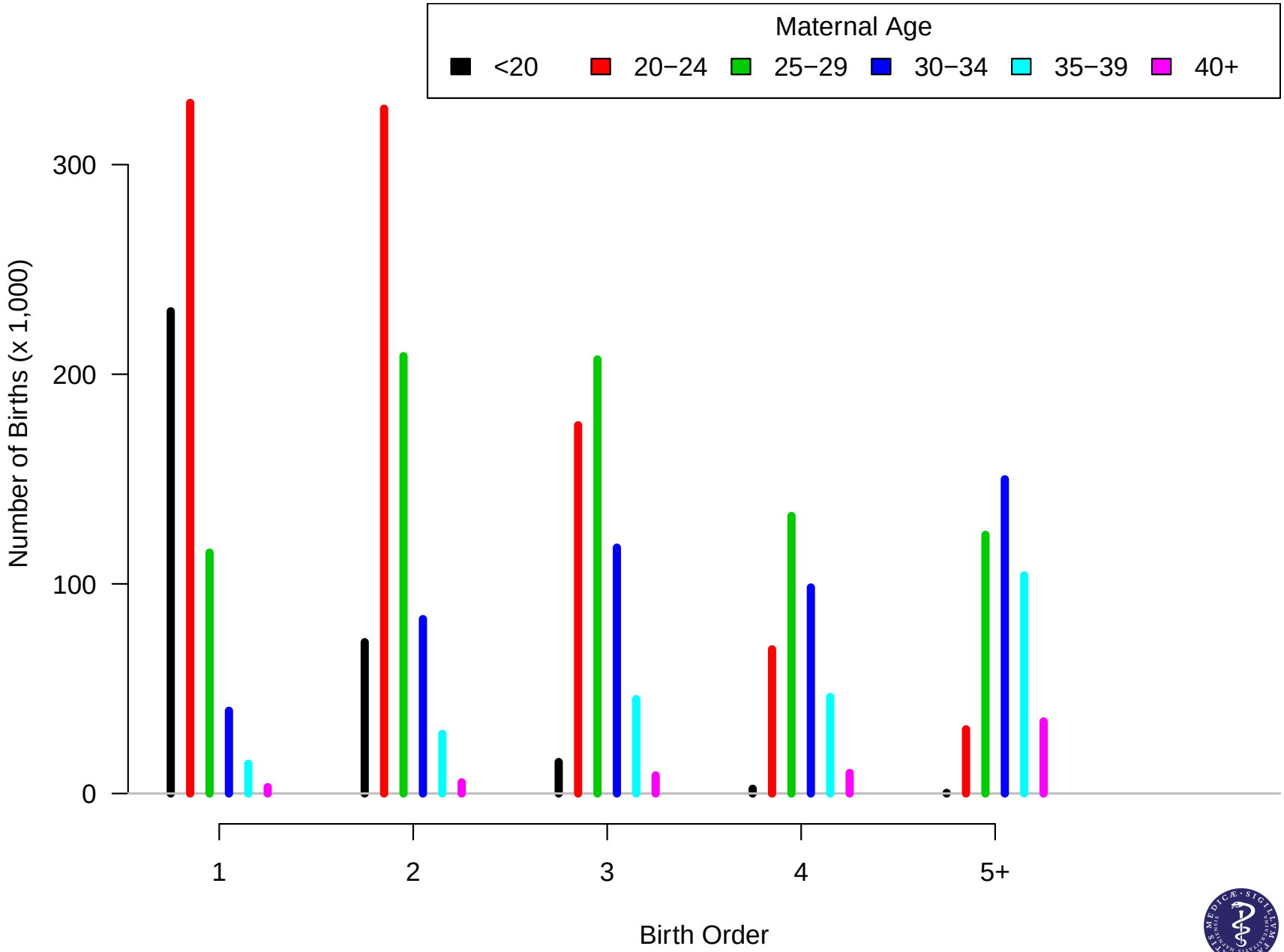
[16] Stark and Mantel (1966) 'Effects of maternal age and birth order on the risk of mongolism and leukemia' J Natl Cancer Inst37(5) 687–698.

Confounding example (birth order and risk of mongolism [16])





[16] Stark and Mantel (1966) 'Effects of maternal age and birth order on the risk of mongolism and leukemia' J Natl Cancer Inst37(5) 687–698.


Maternal Age
<20
20−24
25−29
30−34
35−39
40+
Number of Births (x 1,000)
300
200
100
0
1
2
3
4
5+
Birth Order

When can association mean causation? (1/2)

We usually say that (statistical) association does not imply causation.

Confounding (among others things) can lead to non-causal associations.

There are however some exceptions, i.e., it exists some situations in which an association can be interpreted as a causal association. An example is the situation in which we analyze data from an "ideal" randomized experiments.

When can association mean causation? (2/2)

This is because the randomization ensures that the two compared groups are similar with respect to everything except the intervention / treatment under study. Hence, if a difference in outcome is observed between the two groups, then we can be confident that this is the consequence of this unique difference in exposure / treatment.

In non-randomized (or non "ideally" randomized) experiments the two compared groups will usually differ with respect to more than one characteristic. This generates multiple plausible explanations for the observation of the difference in outcome – some causal and some non causal.

Adjusted analysis

Suppose that in addition to the outcome and the exposure group a categorical confounder variable (e.g. gender) is measured for each individual.

- Subgroup analysis

Analyze 2x2 contingency tables separately in each strata defined by the confounder variable.

- Logistic regression (see Lecture 6)

To compute a "weighted" average of the subgroup analyses, assuming that the exposure-outcome association is the same in all subgroups.[17].

[17] Applicable also with continuous confounders.

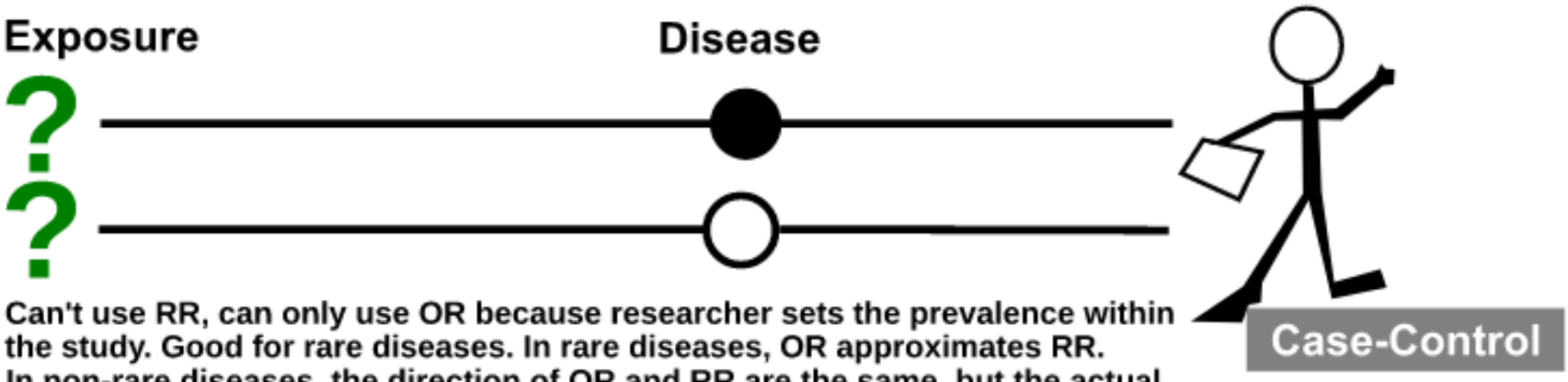
Outline

Observational study design

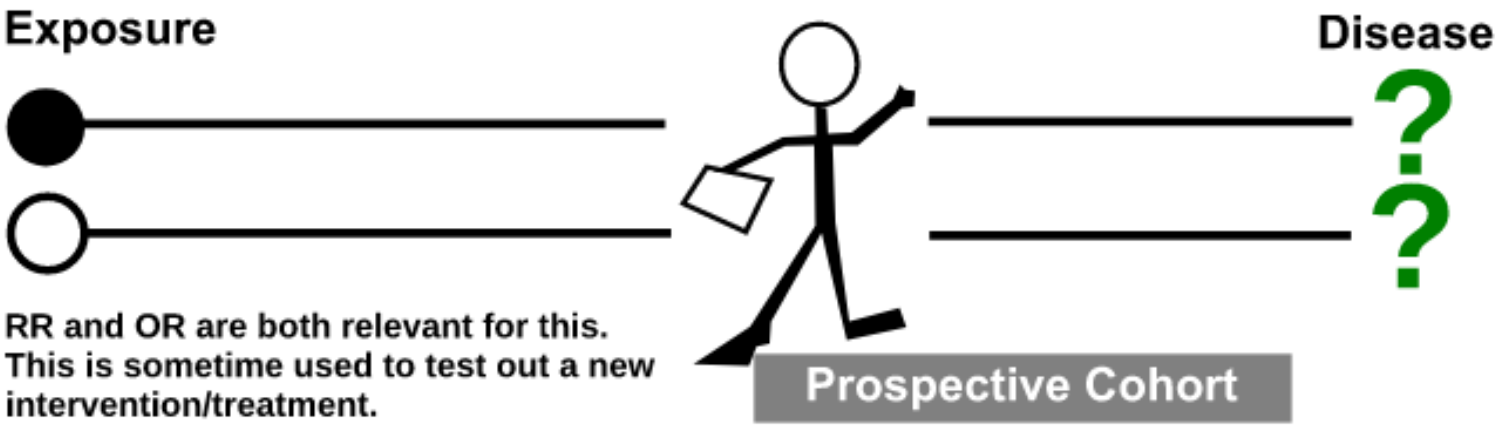
In a prospective **cohort study**, an outcome or disease-free study population is first identified by an exposure (e.g., onset of diabetes) or other inclusion criteria and followed in time until the disease or outcome of interest occurs.

Case-control studies identify subjects by outcome status at the outset of the investigation. First, subjects with outcome are identified and classified as cases. For each case a given number of controls (e.g., 4) are selected. A candidate control is a subject without the outcome but from the same source population.

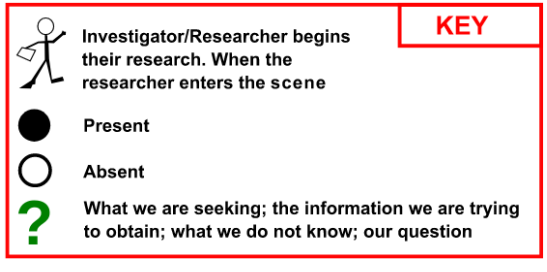
Observational Study Designs: Case Control vs Cohort



Can't use RR, can only use OR because researcher sets the prevalence within the study. Good for rare diseases. In rare diseases, OR approximates RR. In non-rare diseases, the direction of OR and RR are the same, but the actual number obtained of OR and RR are different. You CANNOT obtain a RR for this. It makes no sense to.



RR and OR are both relevant for this. This is sometime used to test out a new intervention/treatment.



source: wikipedia article about case-control studies.

Cohort study: example

For example consider the study from Egerup et al. (2020) [18]

		Infection within first year of life		
		yes	no	total
Kidney-transplanted mother	yes	26	98	124
	no	133	1098	1231
	total	159	1196	1355

The estimated risk ratio is $\widehat{RR} = 1.94$ ($\text{CI}_{95\%} = [1.33; 2.83]$).

[18] Egerup et al. "Increased risk of neonatal complications and infections in children of kidney-transplanted women: A nationwide controlled cohort study." American Journal of Transplantation (2020).

Case-control study: example

For example consider the case-control study of Frachon et al.[19] [20]

		"explained" cardiac problem: yes	no	total
Benfluorex use	yes	19	3	24
	no	8	51	59
	total	27	54	81

$$\widehat{OR} = 40.4$$
$$(\text{CI}_{95\%} : [9.7; 168])$$

The number of controls (here 2 per case) is defined by the study design. Hence we cannot estimate risks as one minus the proportions of controls among exposed and non-exposed...

- The statistic $\widehat{RR}$ depends also on the ratio between controls and cases and should **not** be used for measuring association in case-control studies.
- The statistic $\widehat{OR}$ works.

[19] Frachon et al. "Benfluorex and unexplained valvular heart disease: a case-control study." PLoS one 5.4 (2010).

[20] See movie "150 Milligrams" (2016)

Why does $\widehat{OR}$ work? (1/2)

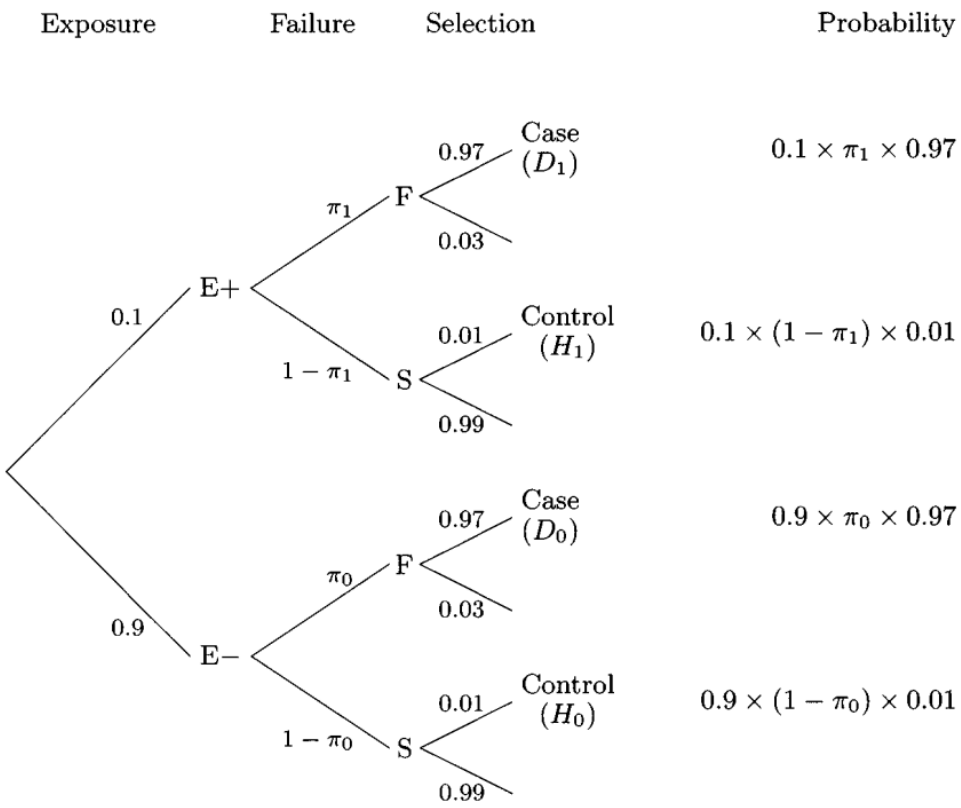


Fig. 16.1. The probability model in the study base.

$$OR = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$$

- 97% of the cases are included in the case-control study and 1% of the "non cases" are selected as controls; all included "blinded" from exposure (i.e. before looking for the information on the exposure).
- Connection to notations of previous slides $\pi_1 = p_1$ and $\pi_0 = p_2$.
- E="exposure", F="Fail", S="Survive", D="Disease", H="Healthy".
- source: "Statistical models in Epidemiology", by Clayton and Hills, page 155.

Why does $\widehat{OR}$ work? (2/2)

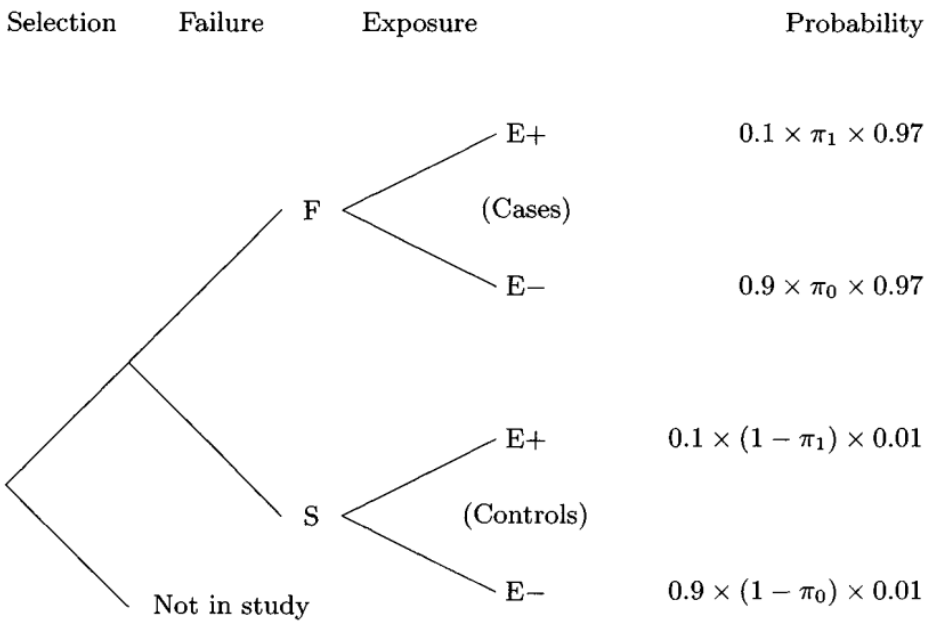


Fig. 16.2. The probability tree for the retrospective argument.

$$\widehat{OR} \approx \frac{\frac{0.1\times\pi_1\times 0.97}{0.1\times(1-\pi_1)\times 0.01}}{\frac{0.9\times\pi_0\times 0.97}{0.9\times(1-\pi_0)\times 0.01}} = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$$

- source: "Statistical models in Epidemiology", by Clayton and Hills, page 156.

Why does $\widehat{OR}$ work? (2/2)

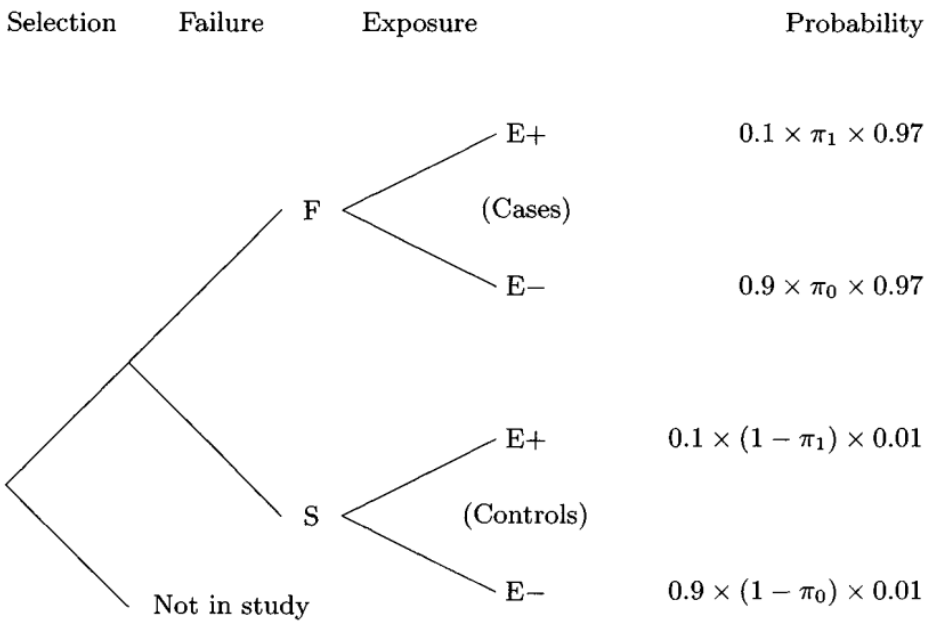


Fig. 16.2. The probability tree for the retrospective argument.

$$\widehat{OR} \approx \frac{\frac{0.1\times\pi_1\times0.97}{0.1\times(1-\pi_1)\times0.01}}{\frac{0.9\times\pi_0\times0.97}{0.9\times(1-\pi_0)\times0.01}} = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$$

but

$$\widehat{RR} \approx \frac{\frac{0.1\times\pi_1\times0.97}{0.1\times\pi_1\times0.97 \;+\; 0.1\times(1-\pi_1)\times0.01}}{\frac{0.9\times\pi_0\times0.97}{0.9\times\pi_0\times0.97 \;+\; 0.9\times(1-\pi_0)\times0.01}} = \frac{\pi_1/\big(\pi_1\times0.97+(1-\pi_1)\times0.01\big)}{\pi_0/\big(\pi_0\times0.97+(1-\pi_0)\times0.01\big)} \neq \frac{\pi_1}{\pi_0}$$

- source: "Statistical models in Epidemiology", by Clayton and Hills, page 156.

Outline

Medical test / screening: jargon

Y: Outcome (disease status) E.g. prostate cancer

X: Test result (biomarker). E.g. $X = \begin{cases} 1 & \text{positive if PSA} > 4.0\,\text{ng/mL} \\ 0 & \text{negative if PSA} \leq 4.0\,\text{ng/mL} \end{cases}$

	$Y = 1$	$Y = 0$
$X = 1$	True positive	False positive
$X = 0$	False negative	True negative

- True positive rate (sensitivity): $P(X = 1 \mid Y = 1)$
- True negative rate (specificity): $P(X = 0 \mid Y = 0)$
- False positive rate (1-specificity): $P(X = 1 \mid Y = 0)$
- The positive predictive value: $P(Y = 1 \mid X = 1)$
- The negative predictive value: $P(Y = 0 \mid X = 0)$