Stats Central

Monthly seminar series

The odd thing about odds ratios

Mark Donoghoe

10 October 2019

Question

The odds ratio of some event in Group 1 vs Group 0 is 0.50.

The probability of the event in Group 0 is 40%.

What is the probability of the event in Group 1?

a. 20%
b. 80%
c. Something else
d. Too hard to figure out in my head

Definition

Probability / risk

- Assume we have a **binary outcome** Y.
- Y can only take one of two possible values, e.g.:
  - Yes / No
  - Dead / Alive
  - Adverse event / No adverse event
  - Response to treatment / No response
  - 1 / 0
- We are often interested in the ***probability*** (or ***risk***) of the "event of interest":

$$p = \Pr(Y = 1)$$

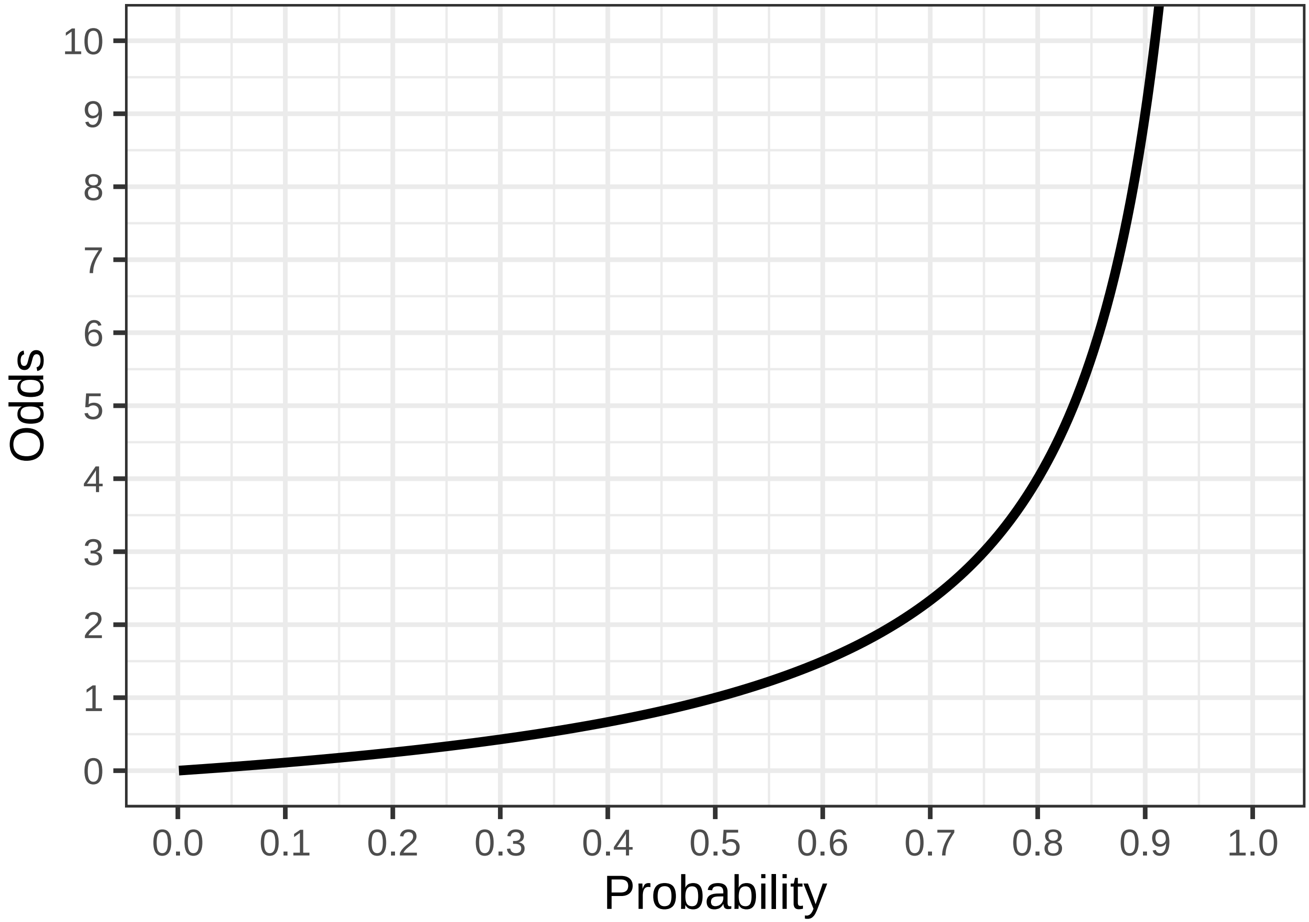
Odds

- The ***odds*** of the event of interest is defined as

$$\text{odds} = \frac{\Pr(Y = 1)}{\Pr(Y = 0)} = \frac{p}{1 - p}$$

- Related to gambling odds: what you win if the event occurs vs what you lose if it doesn't.

Odds vs probability



Odds ratio

- The odds ratio is a **measure of effect**: a way of quantifying the change in probability associated with a change in some variable x, e.g.:
  - Individual characteristic
  - Treatment or intervention
- We will focus on changes in binary variables, e.g.:
  - University degree / No university degree
  - After law change / Before law change
  - Received new treatment / Received placebo
  - 1 / 0
- (But it can also be used for other types of variables)

Odds ratio

- The ***odds ratio (OR)*** is the ratio of the odds of the event of interest in one group ($x = 1$; odds_1) to the odds of the event of interest in the other group ($x = 0$; odds_0).

$$\text{OR} = \frac{\text{odds}_1}{\text{odds}_0} = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} = \frac{p_1(1-p_0)}{p_0(1-p_1)}$$

where $p_j = \Pr(Y = 1 \mid x = j)$.

Calculation and interpretation

Calculation

- If you have estimates of p_0 and p_1, you can plug them into the formula directly:

$$\widehat{\mathrm{OR}} = \frac{\hat{p}_1(1-\hat{p}_0)}{\hat{p}_0(1-\hat{p}_1)}$$

- The Mantel–Haenszel method can be used to combine odds ratios across categorical strata
  - e.g. meta-analysis
- The parameter estimates from a logistic regression model are log odds ratios.
  - Exponentiate to obtain odds ratios.

Interpretation

- If **OR = 1**, the probability of the event in group 1 is **the same** as the probability in group 0.
- If **OR < 1**, the probability of the event in group 1 is **lower** than the probability in group 0.
- If **OR > 1**, the probability of the event in group 1 is **higher** than the probability in group 0.

Interpretation

- Interpreting the number itself (e.g. OR $= 0.5$) is straightforward...
  - *The odds of the event of interest are half as large in group 1 as they are in group 0.*

- ...but not intuitive
  - $p_0 = 0.4$
  - OR $= 0.5$
  - $p_1 =$???

- *"Clinicians are unlikely to find any important question which is answered directly by the odds ratio."*
  - Sinclair, J. C. and M. B. Bracken (1994). J Clin Epidemiol **47**(8): 881-889.

Issues

Difficulty in interpretation

- *"The size of difference, not just the fact of difference, will have direct implications for clinical practice. The measures of treatment effect which are reported should facilitate their practical application."*
  - Sinclair, J. C. and M. B. Bracken (1994). *J Clin Epidemiol* **47**(8): 881–889.

- People typically deal in probabilities rather than odds, so odds ratios are often interpreted as if they are ***relative risks (RR)***:

$$\text{RR} = \frac{p_1}{p_0}$$

UNSW SYDNEY

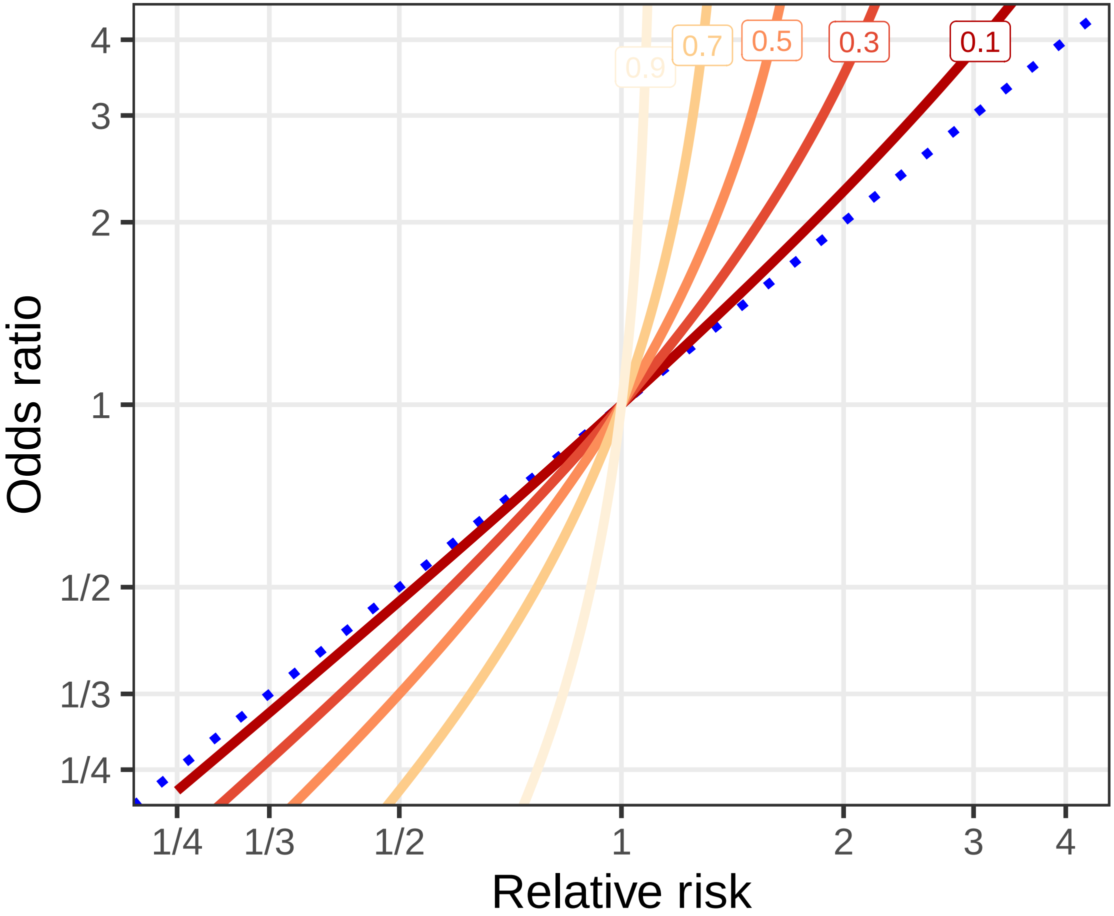
STATS CENTRAL

Difficulty in interpretation

- *"Patients with preoperative haemoglobin levels below 8 g/dl were* ***16.2 times more likely*** *to die than were patients with higher haemoglobin levels."*
  - Carson, J. L. et al (1988). *The Lancet* **331**(8588): 727–729.
  - Relative risk = 8.6.

- *"A* ***threefold higher rate*** *of caesarean section was found in the low-dose oxytocin group* ***(relative risk 2.97)****."*
  - Xenakis E. M-J. et al (1995). *Am J Obstet Gynecol* **173**(6): 1874–1878.
  - Calculated odds ratios but labelled and interpreted as relative risks.
  - Relative risk = 2.47.

Always more extreme



Always more extreme

- The odds ratio is always more extreme (further away from 1) than the relative risk.
- This effect is exaggerated as the baseline risk p_0 increases.
- *"Treating an OR as if it were an accurate estimate of the RR will overestimate both the likely benefits and harms of treatment, and this distortion becomes greater as the disease being treated becomes more severe."*
  - Sackett, D. L., J. J. Deeks and D. G. Altman (1996). *Evid Based Med* **1**(6): 164–166.

Non-collapsibility

- An effect measure is ***collapsible*** if the (weighted) average of stratum-specific effects is equal to the marginal effect.
- This means that a crude (unadjusted) effect will not change if we adjust for a variable that is not a confounder.

- The relative risk is collapsible.
- The odds ratio is **not collapsible**.
  - It does not estimate either the ratio of average odds or the average of stratum-specific odds ratios.
  - Adjusting for more non-confounders will move the OR further away from 1.

- Cummings, P. (2009). *Arch Pediatr Adolesc Med* **163**(5): 438–445.

Advantages

Symmetry

- If we switch events and non-events, i.e. focus on the probability of a non-event. . .

$$q_j = \Pr(Y = 0 \mid x = j) = 1 - p_j$$

- . . . the odds ratio for a non-event will be the reciprocal of that for the event

$$\mathrm{OR}_q = \frac{q_1(1-q_0)}{q_0(1-q_1)} = \frac{(1-p_1)p_0}{(1-p_0)p_1} = \frac{1}{\mathrm{OR}_p}$$

- So it doesn't matter which outcome we choose as the "event of interest".

Symmetry

- The relative risk is **not symmetric**:

$$\text{RR}_q \neq \frac{1}{\text{RR}_p}$$

- *"Any confusion is avoided by adhering to the convention of reporting outcomes as unfavorable (rather than favorable) events, in which case a risk ratio less than unity always signifies a reduction in unfavorable events. . . . We do not argue that [the symmetry of the odds ratio] constitutes a clinically useful advantage."*
  - Sinclair, J. C. and M. B. Bracken (1994). *J Clin Epidemiol* **47**(8): 881–889.

Rare events

- When events are rare (p_0 and p_1 are close to 0), the odds ratio is close to the relative risk.

$$\text{OR} = \frac{p_1(1-p_0)}{p_0(1-p_1)} \approx \frac{p_1}{p_0} = \text{RR}$$

Case-control studies & meta-analyses

- In a case-control study, we cannot estimate the risk, and hence the relative risk is not estimable.
- The odds ratio is estimable using the observed numbers of events.
- The Mantel–Haenszel method can be used to combine ORs across strata, e.g., studies in a meta-analysis.
- Meta-analyses can include case-control and observational designs.

Unconstrained

- The relative risk is constrained by the baseline risk p_0.
  - e.g. if $p_0 = 0.5$, the RR cannot be larger than 2 (otherwise $p_1 > 1$).
- The odds ratio can take any value in $(0, \infty)$.
- Some have claimed this means it is more likely that the OR is constant across individuals or subgroups.
  - It is not possible for both OR and RR to be constant (except OR = RR = 1).
  - $\Rightarrow$ interactions.
- Others have disputed this claim, and/or its relevance.

Logistic regression

- The logit link is the canonical link for the binomial GLM. This provides some nice mathematical properties.
- There are no constraints on the regression parameters.
  - Estimation is more stable.
  - Extrapolation is possible *(but maybe not sensible)*.

Suggestions

Estimation of relative risks

- Relative risks can be estimated in a binomial GLM by using a **log link** function.

```
mdl <- glm(y ~ x1 + x2, data = d,
           family = binomial(link = log))
```

- The parameter estimates are log relative risks.
  - Exponentiate to obtain relative risks.

(Shameless plug)

- Standard estimation methods can run into issues.
- The R package `logbin` provides several estimation algorithms, with greater stability.

Journal of Statistical Software

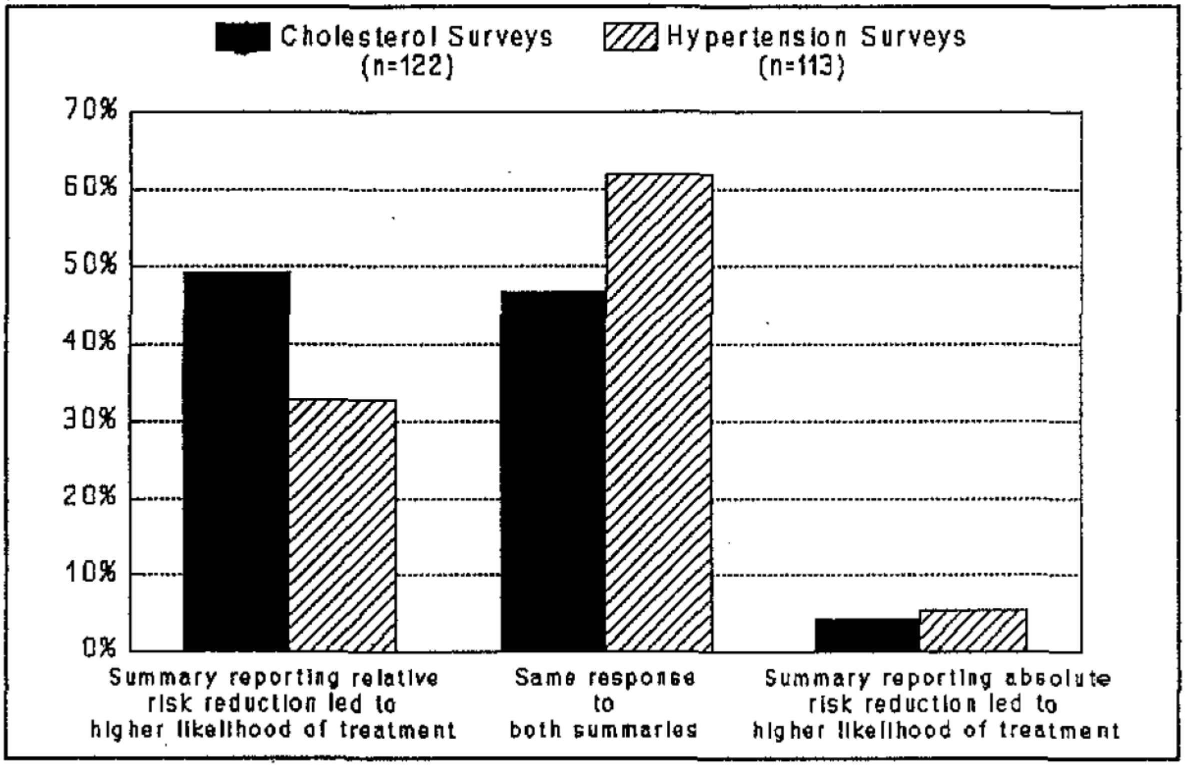
August 2018, Volume 86, Issue 9. *doi: 10.18637/jss.v086.i09*

logbin: An R Package for Relative Risk Regression Using the Log-Binomial Model

Mark W. Donoghoe
Stats Central

Ian C. Marschner
Macquarie University
NHMRC Clinical Trials Centre

Absolute and relative changes



- Forrow, L. F., W. C. Taylor and R. M. Arnold (1992). *Am J Med* **92**(2): 121–124.

CONSORT Statement 2010

- Item 17b:

  For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

- Schultz, K. F., D. G. Altman and D. Moher (2010). *BMJ* **340**(7748): c332

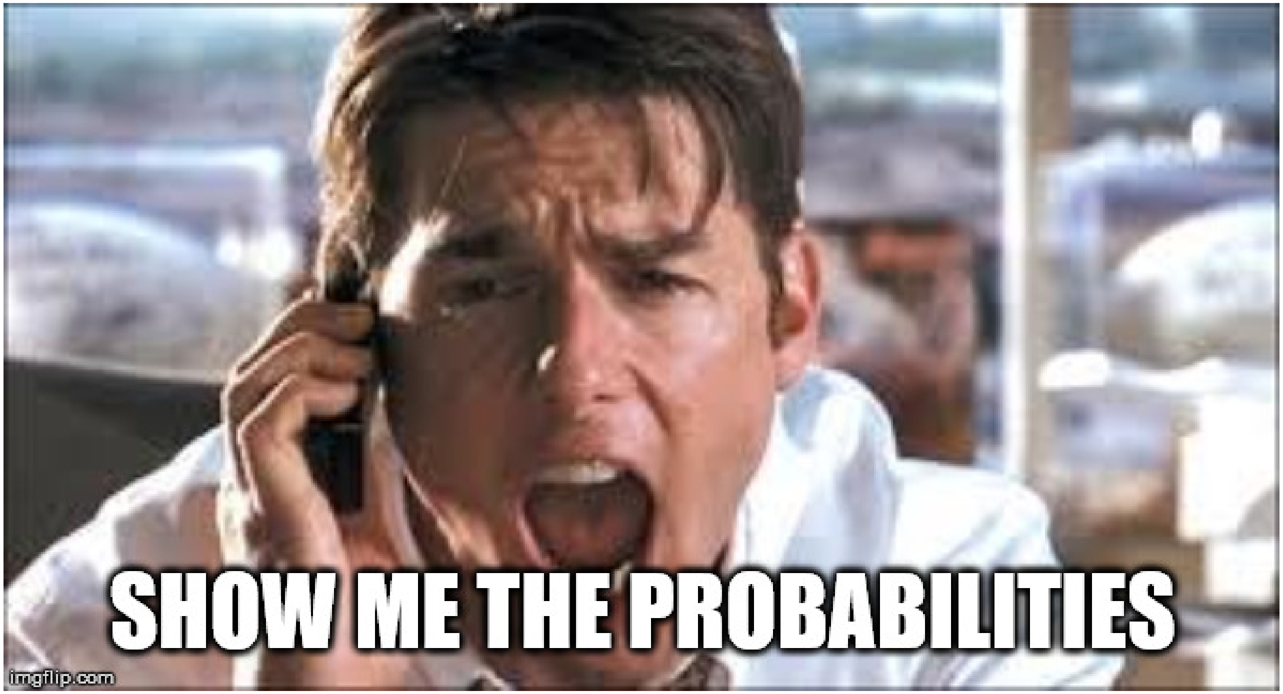
Estimation of risk differences

- Risk differences can be estimated in a binomial GLM by using an **identity link** function

```
mdl <- glm(y ~ x1 + x2, data = d,
           family = binomial(link = "identity"))
```

- The parameter estimates are risk differences.
- (Shameless plug #2): R package `addreg` provides a stable algorithm for fitting this model.

Either way...



Present something meaningful

- *“The odds ratio for the effect of X was 1.33 (95% CI: . . .)”*
- *"This corresponded to an increase in the average risk of Y from. . .*
  - *. . . 1% to 1.3%?*
  - *. . . 23% to 28%?*
  - *. . . 65% to 71%?*

- Confidence intervals
- Subgroups?

Fitted probabilities

- Get fitted probabilities for your data from a model using `predict(mdl, type = "response")`.
- Predict fitted probabilities for new (or hypothetical) data using `predict(mdl, newdata = ..., type = "response")`.
- `emmeans` (formerly `lsmeans`) provides some handy functions for estimating marginal probabilities and effect measures.

Thanks