Biostatistics (MATH11230)
Measures of disease-exposure association

Vanda Inácio

University of Edinburgh

Semester 1, 2022/2023

Measures of disease-exposure association

- We will now be looking at measures that allow us to study the **association between a risk factor or exposure and the occurrence of disease**.

- For simplicity, we will start with a risk factor that can only take two values (e.g., exposed and unexposed).

- Let us denote the disease outcome (also binary: present or absent) by D and the risk factor by E.

Measures of disease-exposure association

Relative risk

- The **relative risk** (RR), also known as risk ratio, for an outcome D associated with a binary risk factor E is defined as

$$\mathrm{RR} = \frac{\Pr(D \mid E)}{\Pr(D \mid \bar{E})} = \frac{\Pr(D \mid E)}{\Pr(D \mid \text{not } E)}.$$

- It is immediate to conclude that the relative risk is a non-negative number.

- A $\mathrm{RR} = 1$, the so-called null value, implies that $\Pr(D \mid E) = \Pr(D \mid \text{not } E)$, which is equivalent to saying that D and E are independent (no association between the exposure to the risk factor E and the outcome D).

- A $\mathrm{RR} > 1$ indicates that there is a greater risk or probability of D when exposed than when not exposed (positive association between the exposure to the risk factor E and the outcome D).

- When $\mathrm{RR} < 1$ there is a reduced risk or probability of D when exposed than when not exposed (negative association between the exposure to the risk factor E and the outcome D).

Measures of disease-exposure association

Relative risk

- We shall note that the relative risk has an implicit upper bound.

- This is because the maximum possible value for a risk and, in particular for $\Pr(D \mid E)$ is one, and therefore $\text{RR} \leq \frac{1}{\Pr(D|\text{not } E)}$.

- This restriction on the range of the relative risk is only problematic if the disease outcome is common, as in this case $\Pr(D \mid \text{not } E)$ may become large.

Measures of disease-exposure association

Relative risk

- We shall note that the relative risk is not symmetric in the role of the two factors D and E.

- That is, the relative risk for E associated with D is a different measure of association than the relative risk for D associated with E, i.e.,

$$\frac{\Pr(E \mid D)}{\Pr(E \mid \text{not } D)} \neq \frac{\Pr(D \mid E)}{\Pr(D \mid \text{not } E)}.$$

Measures of disease-exposure association

Relative risk

- Let us illustrate the relative risk calculation with data from the table below (Jewell, 2003, p 22), which lists the vital status of all births in the United States in 1991, one year after date of birth, and categorised by the marital status of the mother at the birth and the birth weight of the infant.

- Low birth weight was defined as a birth where the newborn's weight is less than 2.5kg.

	Mother's Marital Status		
Infant Mortality	Unmarried	Married	Total
Death	16,712	18,784	35,496
Live at 1 year	1,197,142	2,878,421	4,075,563
Total	1,213,854	2,897,205	4,111,059

	Birth weight		
Infant Mortality	Low Birth weight	Normal Birth weight	Total
Death	21,054	14,442	35,496
Live at 1 year	271,269	3,804,294	4,075,563
Total	292,323	3,818,736	4,111,059

Source: National Center for Health Statistics.

Measures of disease-exposure association

Relative risk

- Following the definition, the relative risk for infant mortality in the USA in 1991, associated with the mother being unmarried, is

$$\frac{\Pr(\text{death} \mid \text{unmarried mother})}{\Pr(\text{death} \mid \text{married mother})} = \frac{16712/1213854}{18784/2897205} = 2.12.$$

- This result indicates that there is a positive association between infant mortality and mother being unmarried.

- The risk of an infant death with an unmarried mother is about twice the risk of an infant death with a married mother.

Measures of disease-exposure association

Relative risk

- In turn, the relative risk for infant mortality in the USA in 1991, associated with a low birth weight is given by

$$\frac{\Pr(\text{death} \mid \text{low birth weight})}{\Pr(\text{death} \mid \text{normal birth weight})} = \frac{21054/292323}{14442/3818736} = 19.0.$$

- There is thus a much greather effect of birth weight on infant mortality than marital status.

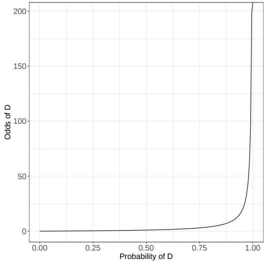
Measures of disease-exposure association

Odds of disease

- The relative risk measures the risk of the outcome D through $\Pr(D)$.
- Another quantity is the **odds** of disease, given by

$$\frac{Pr(D)}{\Pr(\text{not } D)} = \frac{\Pr(D)}{1 - \Pr(D)},$$

  and which gives the likelihood of D occurring relative to it not occurring.
- The odds of D provide the same information as $\Pr(D)$ since knowing one of the quantities immediately determines the other.



Measures of disease-exposure association

Odds ratio

- Just as the relative risk measures association by comparing the probability of the outcome D in the exposed and unexposed groups, the **odds ratio** measures association by comparing the odds of disease in the exposed and unexposed groups.
- The odds ratio for D associated with E is therefore defined by

$$\mathrm{OR} = \frac{\Pr(D \mid E)}{\Pr(\text{not } D \mid E)} \bigg/ \frac{\Pr(D \mid \text{not } E)}{\Pr(\text{not } D \mid \text{not } E)}.$$

- The odds ratio is a number between 0 and ∞.
- Because there is no upper limits for the odds of D, the odds ratio, by opposition to the relative risk, has no implicit upper bound.
- As with the relative risk, $\mathrm{OR} = 1$ is the null value, and it corresponds to the case where the odds of the outcome D are the same in both groups (exposed and unexposed), and it is again equivalent to no association (independence) between D and E.
- When $\mathrm{OR} > 1$, there is a greater risk of D in the exposed group.
- The reverse is true when $\mathrm{OR} < 1$, that is, there is a lower risk of D in the exposed group.

Measures of disease-exposure association

Odds ratio

- $\hookrightarrow$ Let us revisit the example about infant mortality in the USA in 1991.
- $\hookrightarrow$ Let us first compute the odds of infant mortality in the unmarried mother group (exposed group)

$$\frac{16712/1213854}{1197142/1213854} = \frac{16712}{1197142}.$$

- $\hookrightarrow$ Similarly, the odds of infant mortality in the married mother (unexposed) group are

$$\frac{18784/2897205}{2878421/2897205} = \frac{18784}{2878421}.$$

- $\hookrightarrow$ The odds ratio for infant mortality associated with an unmarried mother is

$$\text{OR} = \frac{16712/1197142}{18784/2878421} = 2.14.$$

Measures of disease-exposure association

Odds ratio

- We will now re-do the calcution but now with low birth weight being the risk factor/exposure.

- The odds of death in the low birth weight group are

$$\frac{21054/292323}{271269/292323} = \frac{21054}{271269}.$$

- Analogously, the odds of death in the normal birth weight group are

$$\frac{14442/3818736}{3804294/3818736} = \frac{14442}{3804294}.$$

- Finally, the odds ratio for infant mortality associated with low birth weight is

$$\text{OR} = \frac{21054/271269}{14442/3804294} = 20.4.$$

Measures of disease-exposure association

Odds ratio

↪ So, let us compare what we have obtained for this example so far in terms of relative risks and odds ratios.

	RR	OR
Exposure/risk factor: marital status	2.12	2.14
Exposure/risk factor: low birth weight	19.0	20.4

↪ The RRs and ORs for both exposures (mother's marital status and birth weight) are very similar.

↪ But, is this always the case?

Measures of disease-exposure association

Odds ratio

- Let us look formally at the relationship between the RR and the OR.
- By its definition, the odds ratio for D associated with E is

$$\begin{aligned}
\text{OR} &= \frac{\Pr(D \mid E)}{Pr(\text{not } D \mid E)} \times \frac{\Pr(\text{not } D \mid \text{not } E)}{\Pr(D \mid \text{not } E)} \\
&= \frac{\Pr(D \mid E)}{\Pr(D \mid \text{not } E)} \times \frac{\Pr(\text{not } D \mid \text{not } E)}{Pr(\text{not } D \mid E)} \\
&= \text{RR} \times \frac{\Pr(\text{not } D \mid \text{not } E)}{Pr(\text{not } D \mid E)}.
\end{aligned}$$

Measures of disease-exposure association

Odds ratio

$\hookrightarrow$ Let us suppose first that $\text{RR} > 1$. By definition of relative risk, this means that

$$\Pr(D \mid E) > \Pr(D \mid \text{not } E),$$

which trivially implies that

$$1 - \Pr(D \mid E) < 1 - \Pr(D \mid \text{not } E),$$

and which, in turn, can be equivalently written as,

$$\Pr(\text{not } D \mid E) < \Pr(\text{not } D \mid \text{not } E),$$

thus implying that

$$\frac{\Pr(\text{not } D \mid \text{not } E)}{\Pr(\text{not } D \mid E)} > 1.$$

Measures of disease-exposure association

Odds ratio

- ↪ The OR is thus the product of the RR (which by assumption is greater than one) by a quantity that is also greater than one and therefore we must conclude that OR $>$ RR when RR $>$ 1.
- ↪ When RR $<$ 1, a similar reasoning leads to the conclusion that OR $<$ RR.
- ↪ We thus arrive at the conclusion that the odds ratio is always farther away from one than the relative risk (except when the relative risk is one, in which case the odds ratio will also be one).
- ↪ How farther away the OR is from 1 compared to the RR will depend on both $\Pr(D \mid E)$ and $\Pr(D \mid \text{not } E)$, which are also the two probabilities involved in the computation of the risk ratio.
- ↪ When the risk of disease is low in both the exposed and unexposed groups, then $\Pr(\text{not } D \mid E) \approx 1$ and $\Pr(\text{not } D \mid \text{not } E) \approx 1$ and so OR $\approx$ RR.

Measures of disease-exposure association

Odds ratio

- ↪ We have seen before that the relative risk is not symmetric in the roles of D and E, that is, the relative risk for D associated with E does not (need to) coincide with the relative risk for E associated with D.

- ↪ The odds ratio enjoys the property of being symmetric with respect to the roles of D and E, that is, reversing the roles of D and E makes no difference in its computation.

- ↪ As we will see in one of the next lectures, this property will be key when estimating the association between D and E in certain study designs (case-control studies).

- ↪ In what follows, let $\mathrm{OR}_{D|E}$ denote the odds ratio for D associated with E and $\mathrm{OR}_{E|D}$ denote the odds ratio for E associated with D.

Measures of disease-exposure association

Odds ratio

$$
\begin{aligned}
\mathrm{OR}_{D|E} &= \frac{\Pr(D \mid E)}{\Pr(\text{not } D \mid E)} \Bigg/ \frac{\Pr(D \mid \text{not } E)}{\Pr(\text{not } D \mid \text{not } E)} \\
&= \frac{\Pr(D\&E)/\Pr(E)}{\Pr(\text{not } D\&E)/\Pr(E)} \Bigg/ \frac{\Pr(D\&\text{not } E)/\Pr(\text{not } E)}{\Pr(\text{not } D\&\text{not } E)/\Pr(\text{not } E)} \\
&= \frac{\Pr(D\&E)}{\Pr(\text{not } D\&E)} \Bigg/ \frac{\Pr(D\&\text{not } E)}{\Pr(\text{not } D\&\text{not } E)} \\
&= \frac{\Pr(D\&E)}{\Pr(D\&\text{not } E)} \Bigg/ \frac{\Pr(\text{not } D\&E)}{\Pr(\text{not } D\&\text{not } E)} \\
&= \frac{\Pr(D\&E)/\Pr(D)}{\Pr(D\&\text{not } E)/\Pr(D)} \Bigg/ \frac{\Pr(\text{not } D\&E)/\Pr(\text{not } D)}{\Pr(\text{not } D\&\text{not } E)/\Pr(\text{not } D)} \\
&= \frac{\Pr(E \mid D)}{\Pr(\text{not } E \mid D)} \Bigg/ \frac{\Pr(E \mid \text{not } D)}{\Pr(\text{not } E \mid \text{not } D)} \\
&= \mathrm{OR}_{E|D}
\end{aligned}
$$

Measures of disease-exposure association

Association is not causation!

↪ It is worth finishing this part emphasising that both the relative risk and the odds ratio measure the *association* between the risk factor/exposure and the health outcome.

↪ They do not imply (without further considerations) any causal effect, that is, we cannot conclude that the risk factor causes the health outcome.

↪ Some third factor (known as confounding variable) might be responsible for the relationship between the exposure and the outcome under study (i.e., the groups receiving and not receiving the exposure may be different from one another in some other important variable that is also related to the outcome).

↪ For instance, in the running example of mother's marital status and infant mortality, the apparent association captured by the RR and OR could be due to another factor(s) that is (are) related both to marital status and infant mortality.

↪ We will learn how to deal with confounding variables in a couple of weeks.

Extra material: vaccine efficacy (and its connection to the relative risk...)

- As can be read here:

  `https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(21)00069-0/fulltext`

  vaccine efficacy is generally reported as a relative risk reduction (RRR), which is defined as $1 - \text{RR}$.

- Here the exposure is beneficial and the exposed group consists of those who were vaccinated. The unexposed group then consists of those who were not vaccinated. The outcome D is the getting infected with the virus SARS-CoV-2.

- This blog entry nicely illustrates the calculation of the efficacy of the Pfizer and Moderna (for COVID-19) vaccines

  `https://towardsdatascience.com/pfizer-and-moderna-vaccine-efficacy-calculated-from-data-9566897173c`