Chapter 8. Random Error and the Role of Statistics

December 5, 2024
Epidemiology (10)

Minato Nakazawa
<minato-nakazawa@people.kobe-u.ac.jp>

The roles of statistics in epidemiologic study

- 1st role = Measuring variability to assess the role of chance
  → Dual meanings in the word "*chance*"
  - Outcome of a random process = Unpredictable under any circumstances
    - The coin-toss is usually considered as such random process, the result is completely unpredictable
  - Not easily predictable but are not necessarily random phenomena
    - If we have sufficient information about the initial conditions and the forces applied to the coin, the result of flipping coin can be predicted (Since usually we don't have such information, it's considered as randomizing event). Some individuals can predict flipping coins after practices enough to predict the result
    - If we learn more about the source of error in a body of data, we can reduce errors that may appear random at first ("Physicists tell us that we will never be able to explain all components of error" means Heisenberg's uncertainty principle [see, https://www.youtube.com/watch?v=TQKELOE9eY4]

The roles of statistics in epidemiologic study

- 2nd role = Estimating effects after correcting for biases such as confounding
- Among those two roles, this chapter address the former.
  - Random error is variability. In cohort studies, some of the variability in the outcome is random error, much of variation may reflect hidden biases, unmeasured, even undiscovered.
  - The latter is addressed in Chapter 10 and 12.

Precision of estimation

- The analysis of data should report the magnitude of that epidemiologic quantity and portray the degree of precision with which it is measured
  - (eg.) Case-control study may be undertaken to estimate the IRR between the use of cellular phone and the occurrence of brain cancer
    - The report should present a clear estimate of IRR such as IRR=2.5 – such estimate shown as single value = “point estimate”
    - To show precision, a “confidence interval” (a range of values around the point estimate) is used. → Wide confidence interval = low precision, narrow confidence interval = high precision

POINT ESTIMATES, CONFIDENCE INTERVALS

- Point estimate (being a single value) cannot express the statistical variation (random error)
- Large study → estimation process is comparatively precise → little random error
- Small study → less precision, more random error
- Confidence interval: The amount of random error in the estimate
  - A given confidence interval is tied to an arbitrarily set level of confidence (eg., 95%, 90%, …)
  - Statistical definition: The level of confidence is set to 95% → If the data collection and analysis could be replicated many times and the study were free of bias, the confidence interval would include within it the correct value of the measure 95% of the time.
  - The statistical definition presumes (1) the difference among replication is only caused by chance, (2) variability in the data is adequately described by a statistical model without any bias (or bias is completely controlled) → unrealistic even in randomized controlled trials
  - Practical meaning: A rough estimate of the statistical variability (actually random error) in a set of data

P VALUES

- The confidence interval is calculated from the same equations that are used to generate another commonly reported statistical measure, the *P* value (the statistic used for statistical hypothesis testing).
- The *P* value is calculated in relation to a specific hypothesis (usually **null hypothesis** = there is no relation between exposure and disease)
  - For RR measure, null hypothesis is RR=1.0
    - For RD measure, null hypothesis is RD=0
  - The *P* value represents **probability** (under the condition that the null hypothesis is true, the study is free of bias, and the observations are independent each other) that the data obtained in the study would demonstrate an association as far from the null hypothesis or farther than what was actually obtained. → The required condition is seldom met, so that the P value is not a meaningful probability. Instead, a measure of relative consistency between the null hypothesis and the data in hand.
  - More specifically, if a *P* value were as small as 0.01, it would mean that the data were not very consistent with the null hypothesis.

STATISTICAL HYPOTHESIS TESTING VERSUS ESTIMATION

- Often a *P* value is used to determine the presence or absence of statistical significance. Statistical significance is sometimes considered as meaningful term, but only means whether the *P* value is less than some arbitrary value (almost always 0.05).
- Statistical hypothesis testing is used to describe the process of deciding whether to reject or not to reject a specific hypothesis (usually null hypothesis).
  - If an analysis gives a result of statistically significant, the null hypothesis is rejected as false.
  - If a result is not statistically significant, it means that the null hypothesis cannot be rejected. It doesn't mean that the null hypothesis is correct.
  - The declaration of "statistically significant" or "not significant" is less informative than giving a *P* value.
  - Separation of the information on strength of relation and precision is important. Thus both point estimates and precision information (*P* value or confidence interval) are needed.

P-value functions = Confidence interval functions: See R code, https://minato.sip21c.org/epispecial/codes-for-Chapter8.R

- *P*-value function (= confidence interval function) enlarges on the concept of the *P* value.
- Instead of testing only the null hypothesis, we can also calculate a *P* value for a range of other hypothesis.
  - Consider RR ranging from 0 to infinity, equaling to 1.0 if the null hypothesis is correct.
  - We can calculate infinite number of *P* values that test every possible value of RR.
  - If we do so and plot the result, it becomes *P*-value function.
- (eg. Figure 8-1) The ordinary *P* value is 0.08. Point estimate of OR is 3.2 (In main text, it's written a RR, but actually OR). Though *P* value is larger than 0.05, *P*-value function in Figure 8-1 makes it clear that there is a strong association in the data. Since the *P* values are same for RR=1 and RR=10.5, there is no reason to prefer the interpretation of RR=1 over the interpretation of RR=10.5.

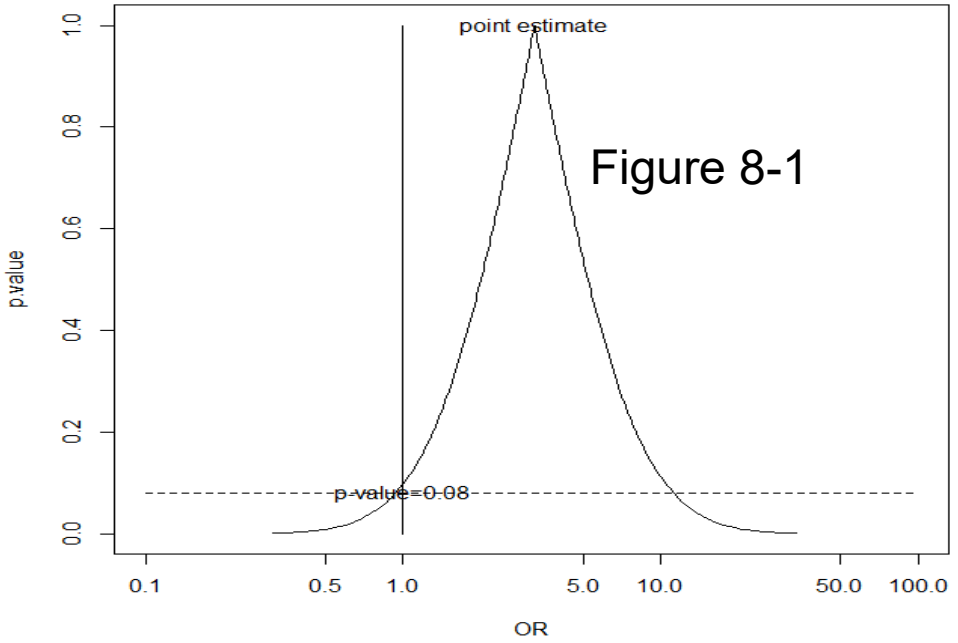


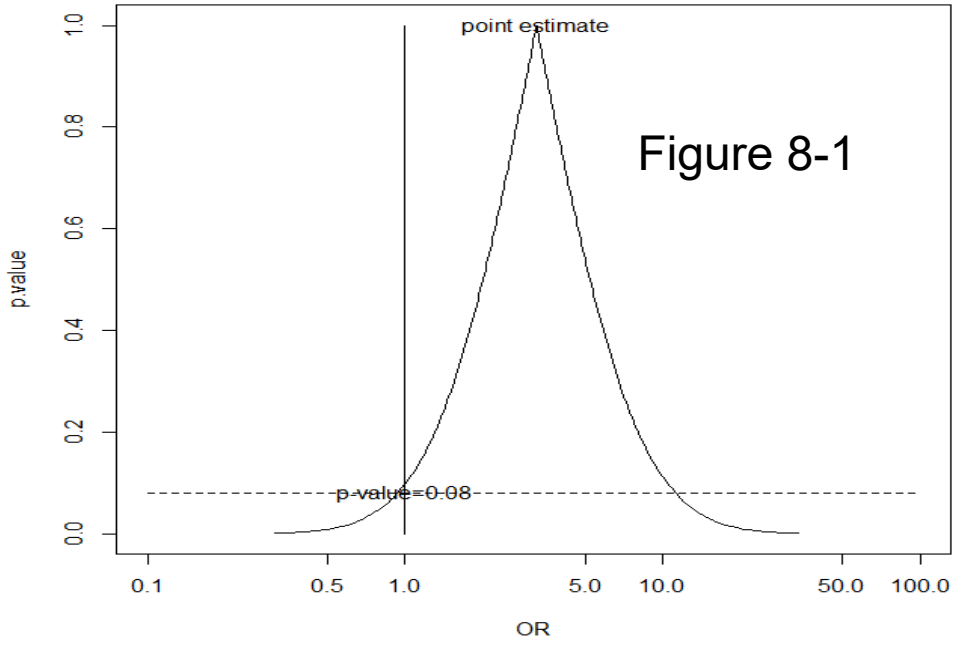
Figure 8-1

Table 8-1. Case-control data for congenital heart disease and chlordiazepoxide use in early pregnancy

	Chlordiazepoxide use		
	Yes	No	Total
Cases	4	386	390
Controls	4	1250	1254
Total	8	1636	1644
OR = (4/386)/(4/1250) = 1250/386 = 3.2			

R code to draw Fig. 8-1

```
if !require(fmsb) { install.packages("fmsb")
 library(fmsb) }
# Figure 8-1
TAB1 <- matrix(c(4, 4, 386, 1250), 2)
T8.1 <- pvalueplot(TAB1, plot.OR=TRUE, plot.log=TRUE,
        xrange=c(0.1, 100))
res <- oddsratio(TAB1)
segments(1, 0, 1, 1, lty=1) # vertical line
segments(0.1, res$p.value, 100, res$p.value, lty=2)
text(res$estimate, 1, "point estimate")
text(1, res$p.value, sprintf("p-value=%3.2f", res$p.value))
```




```
# definition of pvalueplot()
function (XTAB, plot.OR = FALSE, plot.log = FALSE,
          xrange = c(0.01, 5), add = FALSE, ...)
{   x.a <- XTAB[1, 1]
    x.b <- XTAB[1, 2]
    x.c <- XTAB[2, 1]
    x.d <- XTAB[2, 2]
    x.N1 <- sum(XTAB[, 1])
    x.N0 <- sum(XTAB[, 2])
    cp <- c(1:9/1000, 1:9/100, 10:90/100, 0.9+1:9/100, 0.99+1:9/1000)
    cpx <- c(cp, 1, rev(cp))
    cpy <- c(cp/2, 0.5, 0.5 + cp/2)
    cRR <- exp(log(x.a * x.N0/x.b/x.N1) +
          qnorm(cpy) * sqrt(1/x.a - 1/x.N1 + 1/x.b - 1/x.N0))
    cOR <- exp(log(x.a * x.d/x.b/x.c) +
          qnorm(cpy) * sqrt(1/x.a + 1/x.b + 1/x.c + 1/x.d))
    if (plot.OR) { rval <- data.frame(OR = cOR, p.value = cpx) }
    else { rval <- data.frame(RR = cRR, p.value = cpx) }
    OpLog <- ifelse(plot.log, "x", "")
    if (add) { lines(rval, ...) }
    else { plot(rval, type = "l", xlim = xrange, log = OpLog, ...) }
    return(rval) }
```

P-value functions (cont'd)

- Table 8-2 is the data with large sample size.
- Point estimate of OR is 1.1
- A *P* value for null hypothesis is 0.04 (less than 0.05), so "significant".
- The narrowness of the p-value function for Table 8-2 reflects the larger sample size, which only means better precision.
- In Table 8-1, the association is "not significant" but the study is properly interpreted as raising concern about the effect of the exposure. In Table 8-2, the study provides reassurance about the absence of a strong effect, but the significance test gives a result that is "significant".

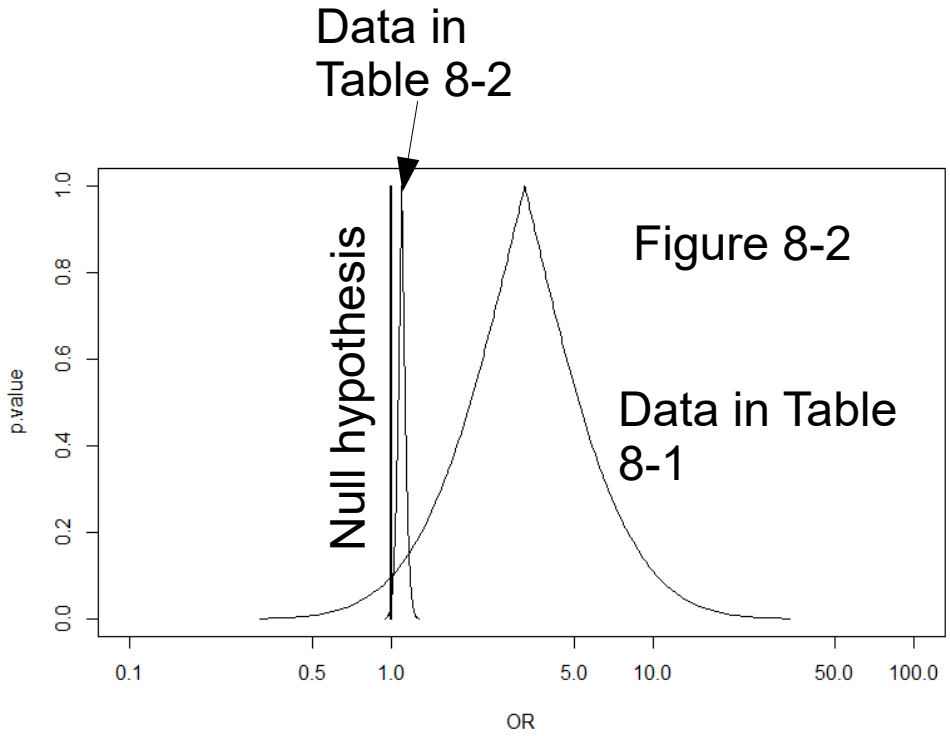


Table 8-2. Hypothetical case-control data			
	Exposure		
	Yes	No	Total
Cases	1090	14910	16000
Controls	1000	15000	16000
Total	2090	29910	32000
OR = (1090/14910)/(1000/15000) = 1.1			

P-value functions (cont'd)

- The most straightforward way to get both messages (strength of association and the effect of precision) is from the upper and lower confidence limits: The two numbers that form the boundaries to a confidence interval.
- The *P*-value function is closely related to the set of all confidence intervals for a given estimate (see, Figure 8-3, showing p-value function as nested confidence intervals).

- Instead of showing full *P*-value function, a single confidence interval is sufficient to determine the entire *P*-value function.
- Confidence intervals are too often not interpreted with the image of the corresponding *P*-value function in mind. Sometimes only used to determine statistical significance by whether the confidence interval includes null hypothesis or not. It's regrettable.

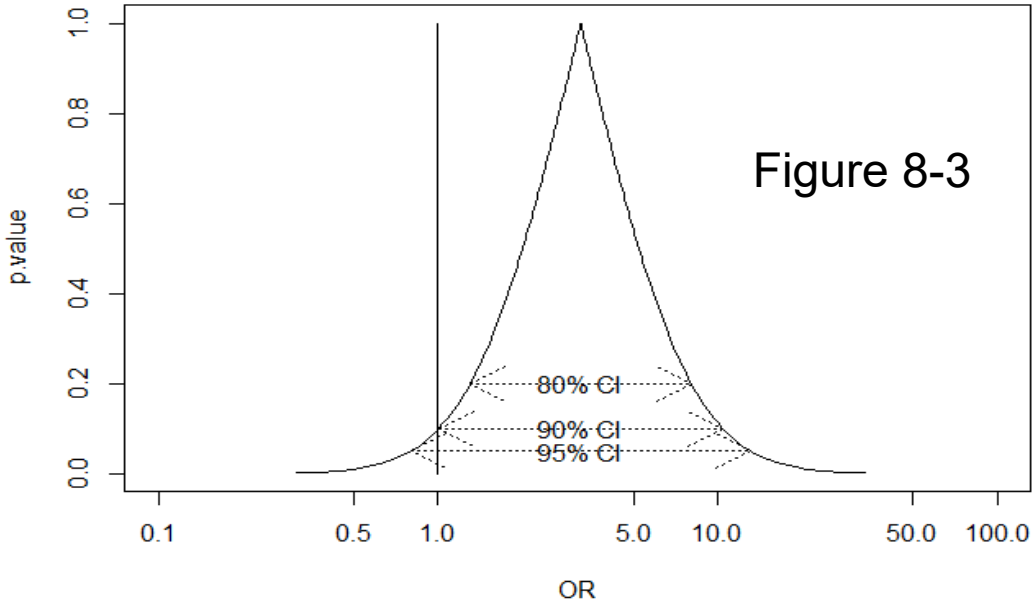


Figure 8-3

Example: Is flutamide effective in treating prostate cancer?

- Figure 8-4 shows the result of meta-analysis (the calculation of pooled odds ratio, OR_{MH} is explained in Chapter 10, p.187)
  - The result of 10 studies showed “not significant” protective effect of flutamide on prostate cancer (OR_{MH}=0.89, p>0.05)
  - By adding Table 8-3 data (the study itself is also “not significant”, OR=0.87, p=0.14), the result of 11 studies showed “significant” protective effect (OR_{MH}=0.89, p<0.05), simply due to the improvement of precision.
  - However, p-value functions are almost same.

- https://www.ncbi.nlm.nih.gov/pubmed/7630245
- https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(05)74403-2/fulltext

Table 8-3. Summary of survival data from the study of flutamide and prostate cancer

	Flutamide	Placebo
Died	468	480
Survived	229	205
Total	697	685
OR = 0.87, 95%CI = 0.70-1.10		

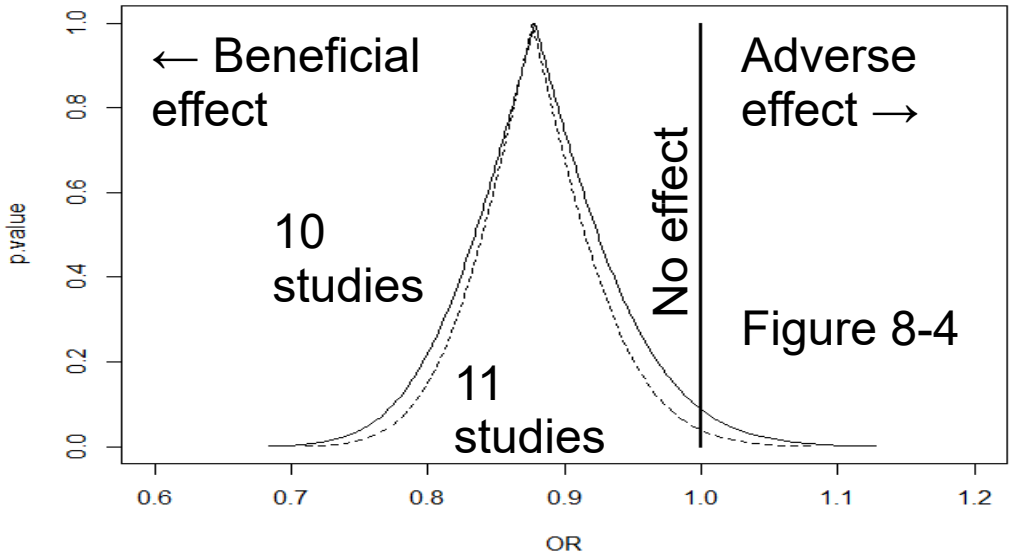


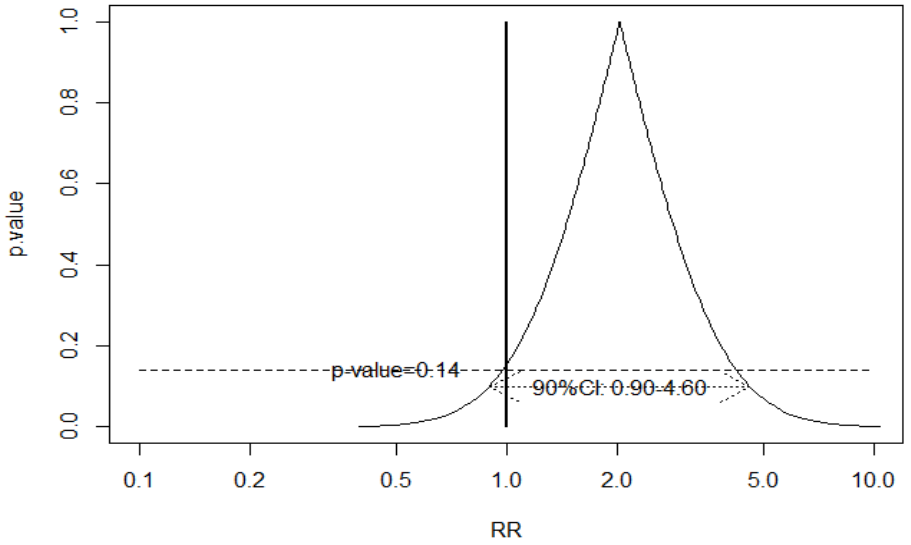
Figure 8-4

Example: Is St. John's wort effective in relieving major depression?

- Extract of St John's wort have long been used as a folk remedy for depression
- Table 8-4 is the result of randomized trial only among those whose depression was relatively less severe
- RR=2.0 means that the use of St. John's wort increased the remission in symptoms two times more than placebo group.
- The authors misinterpreted the result by the lack of statistical significance (p=0.14 for null hypothesis)
- Figure 8-5 shows that the data regarding remissions among the less severely affected patients hardly support the theory that St. John's wort is ineffective.

- https://jamanetwork.com/journals/jama/fullarticle/193754

Table 8-4. Remission among patients with less severe depression		
	St. John's wort	Placebo
Remission	12	5
No remission	47	45
Total	59	50
RR = (12/59)/(5/50) = 2.0, 90%CI: 0.90-4.6		



Simple approaches to calculating confidence intervals

- Confidence intervals are usually calculated on the presumption that the estimate comes from the statistical distribution called a *normal distribution*, the usual bell-shaped curve.
- Estimates based on the normal distribution are always reasonable with enough data. If data are sparse, specialized formula for small numbers (exact methods) are needed.
- A given normal distribution is described with regard to its mean and its spread (SD)

- For rate difference (or risk difference)
  - The formula [8-1] gives 90% confidence interval for the rate difference.
  - RD_L is lower confidence limit
  - RD_U is upper confidence limit
  - 1.645 corresponds to 90% confidence. 1.96 corresponds to 95% confidence
  - In R, qnorm(1-(1-0.90)/2) gives 1.645. The normal distribution from qnorm((1-0.90)/2) to qnorm(1-(1-0.90)/2) includes 90% of data.

$$\mathrm{RD}_L, \mathrm{RD}_U = \mathrm{RD} \pm 1.645 \times \mathrm{SD} \quad [8-1]$$

- For rate ratio (or risk ratio or odds ratio)
  - The distribution of ratio measures is asymmetrically skewed toward large values. To conteract the skewness, it is customary to set the confidence limits on the log scale (after a logarithmic transformation, take exponential again)
  - [8-2] gives 90% confidence interval for the risk ratio

$$\ln(\mathrm{RR}_L), \ln(\mathrm{RR}_U) = \ln(\mathrm{RR}) \pm 1.645 \times \mathrm{SD}(\ln(\mathrm{RR}))$$

$$\mathrm{RR}_L, \mathrm{RR}_U = e^{\ln(\mathrm{RR}) \pm 1.645 \times \mathrm{SD}(\ln(\mathrm{RR}))} \quad [8-2]$$

Statistical significance testing versus estimation (box)

- Significance testing is qualitative, not quantitative
- Ideally, a confidence interval should be viewed as a tool to conjure up an image of the full P-value function, a smooth curve with no boundary on the estimate
- In most instances, there is no need for any test of statistical significance to be calculated, reported, or relied on, and we are much better off without them.