Inference re Epidemiologic Parameter: **Rate or Incidence Density**

Theoretical: *Rate* or ID or λ

Empirical: c cases in PT population-time units : $\widehat{ID} = \hat{\lambda} = c \div PT$;

Model: $C \sim Poisson(\mu)$, where $\mu = \lambda \times PT$; c : realization of C

P-value: $\mathrm{H}_0 :\ ID = ID_0,\ \lambda = \lambda_0;\ \rightarrow \mu = \lambda_0 \times PT = \mu_0$;

<u>Exact</u>

- $P[\ C \le c\ \&\ C \ge c \mid \mu_{null}\ ]$ (lower & upper-tails)

<u>Approx.</u>

- N'l Approx. to distr'n of C or transform, $t(C)$, of C

CI 100(1-α)%: <u>Exact</u>

- $P[C \ge c|\mu_L] = \alpha/2;\ P[C \le c|\mu_U] = \alpha/2; \rightarrow \{\mu_L, \mu_U\} \div PT$

<u>Approx.</u>

- reverse transform of $ci = \{t(c) \mp z_{a/2} \times SE[t(c)]\}; ci \div PT$

transform	$t(c)$	$SE[t(c)] = Var^{1/2}$	CI = reverse of ci
identity	c	$c^{1/2}$	n/a
log	$\log[c]$	$\{1/c\}^{1/2}$	$e^{\{ci\}}$ or $\exp\{ci\}$
sqrt	$\sqrt{c}$	$\{0.25\}^{1/2} = 0.5$	$\{\sqrt{c} \mp 0.5 z_{\alpha/2}\}^2$

[1] **Comparison** of ID's or Rates in *index* ($_1$) *vs. reference* ($_0$) category

[2] Theoretical: $\lambda_1\ \&\ \lambda_0 \rightarrow \lambda_1 - \lambda_0\ (IDD);\quad \lambda_1 \div \lambda_0\ \ (IDR)$

[3] Empirical: $c_1/PT_1\ \&\ c_0/PT_0 \rightarrow \hat{\lambda}_1 - \hat{\lambda}_0;\ \hat{\lambda}_1 \div \hat{\lambda}_0$

[4] Model: $c_i \sim Poisson(\mu_i = \lambda_i \times PT), i = 0, 1$; c_1 independent of c_0.

[5] P-value: $\mathrm{H}_0 :\ IDD = 0;\ IDR = 1;$

[6] <u>Exact</u>

[7] - $c_1|(c_1 + c_0) \sim Binomial(''n'' = c_1 + c_0, \pi = PT_1/\{PT_1 + PT_0\})$

[8] <u>Approx.</u> [using the $\hat{\lambda}_i$'s, or transforms, $t(\hat{\lambda}_i)$, of them]

[9] - $z = \sqrt{X^2} = \{t(\hat{\lambda}_1) - t(\hat{\lambda}_0)\}/\{Var_{H_0}[t(\hat{\lambda}_1)] + Var_{H_0}[t(\hat{\lambda}_0)]\}^{1/2}$

[10] CI: <u>Exact – IDR only</u>

[11] - $IDR_L : P[\ \ge c_1|IDR_L] = \alpha/2;\ IDR_U : -\ similarly$

[12] <u>Approx. – both IDD and IDR:</u> ci on t scale → CI on desired scale

[13] - ci: $\{t(\hat{\lambda_1}) - t(\hat{\lambda_0}) \pm z_{a/2}(Var[t(\hat{\lambda_1})] + Var[t(\hat{\lambda_0})])^{1/2}\} \rightarrow CI$

[14]

measure	transform	$t(\hat{\lambda})$	$ci \rightarrow CI$	test-based* CI
ID <u>D</u>iff.	identity	$\hat{\lambda}$	n/a	$\widehat{IDD} \times (1 \pm z_{\alpha/2}/X)$
ID <u>R</u>atio	log	$\log[\hat{\lambda}]$	e^{ci}	$\widehat{IDR}^{(1 \pm z_{\alpha/2}/X)}$

[15] [16] [17]

<u>Notes</u>:

[1] "Rate" in the *incidence density* sense.

[2-4] helps to separate obs'd & exp'd *numerator*, μ & c, from *rate* $(ID, \lambda, \hat{\lambda})$. $\mu = \lambda \times PT$

[3] We could use the usual 'Y' as the numerator (i.e., the count) but 'C' more meaningful.

[5] Interested in λ, not μ, but it is C which has the Poisson distribution!

[7] `ppois(c,μ)` & `1-ppois(c-1,μ)` in `R`; `Poisson(c,μ,T)` in Excel; See c634 Resources.

[11] via:- Tables; `cii PT c, poisson` in `Stata`; `pois.exact` in `epitools` in `R`; etc.

[11] via trial & error using `ppois` or `Poisson`; or exactly using Poisson ⇔ Chi-sq link.

[11 &13] Note that CI for λ or ID is of the form $\{CI$ for $\mu\} \div PT$.

[13] Using '$t()$' as shorthand for a generic '<u>t</u>ransform' or 'function of'.

[14] These transforms will be called '***links***' when we come to generalized linear models.

[15] The variance of a Poisson random variable is a function *only* of the mean μ.

[15] Rothman2002 ('BabyRothman') p132 uses identity link, i.e. untransformed version.

[16] Log link typically used for rate ratio; makes more sense than identity link for 1 rate.

[17] Variance-stabilizing transformation.

<u>Notes</u>:

[4] Assuming *independent* samples.

[7] [11] Fixing $c_1 + c_0$ eliminates 1 nuisance parameter leaving just the ratio $IDR = \lambda_1/\lambda_0$.

[9] Again here, several *equivalent* versions of X^2 for 2 counts. See jh c607/ch9.

... Don't force c_1, c_0, PT_1, PT_0 into a 2×2 table. See depiction as a '2×1' table (jh Ch9).

[11] Use Binomial distr'n; π is determined by (is function of) the IDR & ratio of the PT's.

... Use def'n. of μ's to show: $\pi = \mu_1/(\mu_1 + \mu_0) = IDR/(IDR + PT_0/PT_1)$

... Lower limit π_L for $\pi \rightarrow$ lower limit for IDR: $\mathrm{IDR}_L = \{\pi_L/(1 - \pi_L)\} \div \{PT_1/PT_0\}$ etc.

... Can use same `Excel` spreadsheet (jh c607 ch 8 resources) for exact *test* and exact *CI*.

... Via `Stata` immediate command `iri` c_1 c_0 PT_1 PT_0 or `rateratio.*` in `epitools`

[14] Can also use a test-based CI for a *risk* difference/ratio or an odds ratio.

... Test-based CI's use the $Variance\ under\ the\ Null$, used when testing the null value.

[17] $Var[\log\{I\hat{D}R\}]$ had just 2 terms, $1/c_1 + 1/c_0$. Since PT_i is just a constant, $Var[log(\hat{\lambda}_i)] = Var[log(c_i/PT_i)] = Var[log(c_i)] + Var[log(PT_i)] = 1/c_i + 0 = 1/c_i$. In contrast, the variance of a *difference* of 2 IDs depends on both the 2 c's *and* the 2 PT's.

[15-17] Rothman2002Ch7 emphasizes ease of manual calculation over heuristics.

Comparison of ID's (Rates, λ's) in index$_{(1)}$ vs. ref$_{(0)}$ categories - **stratified data**

Assuming a single (i.e., summary) Rate Ratio or Rate Difference makes sense.

Empirical: $c_{1,s}$ cases in $PT_{1,s}$ and $c_{0,s}$ cases in $PT_{0,s}$ in stratum 's' ($s = 1, \ldots, S$)

Model: $2S$ indep't rv's $c_{0,s} \ldots c_{1,S}$: $c_{i,s} \sim Poisson(\mu_{i,s})$; $\mu_{i,s} = \lambda_{i,s} \times PT_{i,s}$.

ID **Ratio** (IDR): aliases: Rate Ratio; Incidence Ratio ('IR') - Rothman's term

(1) Antilog of weighted average ($W^{td}Ave.$) of stratum-specific $\log\{\widehat{IDR}\}$'s

Weights $\{w_1, w_2, \ldots w_s, \ldots w_S\}$ are *precision*-based

As in Woolf's formula for combining $\widehat{OR}$s in a c-c study.

N.B.: standardization uses another type of weights (NOT precision-based).

$w_s = \frac{1/V_s}{1/V_1 + \ldots + 1/V_S}$; $V_s = Var[\log\{\widehat{IDR}_s\}] = 1/c_{1,s} + 1/c_{0,s}$

Woolf's variance formula has 2 additional terms; these 2 terms pay for the uncertainty in estimating the PT's using a 'denominator' ('control') series.

Point Estimate: $\widetilde{IDR} = \exp\left[\sum_s w_s \times \log\{\widehat{IDR}_s\}\right] = \exp\left[W^{td}Ave.\right]$

Sampling variation of $\log\left[\widehat{IDR}\right]$'s more Gaussian than $\widehat{IDR}$s themselves.

Variance of $W^{td}Ave$: $Var = 1/\{\sum_s 1/V_s\}$; $SE = Var^{1/2}$

CI: $\exp\left[W^{td}Ave \mp z_{\alpha/2} \times SE\{W^{td}Ave\}\right] = \widetilde{IDR} \div\times \exp\left[z_{\alpha/2} \times SE\right]$

Think of $\exp\left[z_{\alpha/2} \times SE\right]$ as a '***multiplicative***' Margin of Error (M.E.)

Instead of $\hat{\theta}$ **minus/plus** $M.E.$, it's $\hat{\theta}$ **divided/multiplied by** $M.E.$.

(2) Mantel-Haenszel Summary IDR See Rothman2002Ch8

The 1959 MH summary measure was for ORs; for many years, its variance defied statisticians. $Var\Big[\log\left[\widetilde{IDR_{MH}}\right]\Big]$ was less challenging.

No. cases, PT in stratum s : $c_s = c_{1,s} + c_{0,s}$; $PT_s = PT_{1,s} + PT_{0,s}$

Point Estimate: $\widetilde{IDR_{MH}} = \frac{\sum_s\{c_{1,s} \times PT_{0,s}\} \div PT_s}{\sum_s\{c_{0,s} \times PT_{1,s}\} \div PT_s} = \frac{Num_{MH}}{Den_{MH}}$

Mantel's '1 ratio' formulation gives stability – no averaging of S unstable ratios!

Variance of $\log\left[\widetilde{IDR_{MH}}\right]$: $Var = \frac{\sum_s\{c_s \times PT_{1,s} \times PT_{0,s}\} \div PT_s^2}{Num_{MH} \times Den_{MH}}$

Rothman's formulation seems more suited for his 1970s hand calculator.

The formula in Table 8-4 is **incorrect**. For **correct** version see the p156 e.g.

CI: $\exp\left[\log\left[\widetilde{IDR_{MH}}\right] \mp z_{\alpha/2} \times Var^{1/2}\right] = \widetilde{IDR_{MH}} \div\times$ '$M.E.$'

Again, notice the *multiplicative* '$M.E.$': point est. $\div\times$ $M.E.$ instead of $\mp M.E.$

ID **Diff.** (IDD): aliases: Rate Diff.; Incidence Difference ('ID') - Rothman's term

Precision-weighted average of stratum-specific $\widehat{IDD}$'s See Rothman2002Ch8

Standardization uses another type of weights.

$w_s = \frac{Q_s}{Q_1 + \ldots + Q_S}$; $Q_s = 1/(1/PT_{1,s} + 1/PT_{0,s})$; $V_s = \frac{c_{1,s}}{PT_{1,s}^2} + \frac{c_{0,s}}{PT_{0,s}^2}$

Q_s is proportional to the information (inverse of variance) in stratum s.

Point Estimate: $\widetilde{IDD} = \sum_s w_s \times \widehat{IDD}_s$

Again, Rothman's eqn. 8-4 seems designed to minimize calculator keystrokes.

CI: $\widetilde{IDR} \mp z_{\alpha/2} \times SE$; $SE = Var^{1/2}$; $Var = \sum_s w_s^2 \times V_s$.