Inferences regarding a single event rate parameter: i.e. rate of events per N [$=10^x$] units of experience

data: **c "events"** counted in sample of n units of "experience"; or Binomial(c,n) if c << n.

[can use c to calculate a rate i.e. empirical rate = $\frac{c}{n}$ × N events per N units of experience; N usually 10^3 or 10^4 or the like]

See "Modern Epidemiology"(Rothman 1986) ; Observation & Inference (Walker) or Epidemiology: An introduction (Rothman, 2002, 133-134).

	Small no. of events	Large no. of events
CI for μ = E[c] E[c] is a parameter: the theoretical (unobservable) average number of events per n units; c refers to the realization in the observed sample ***Example***: If observe y=2 cases of leukemia in a certain amount of experience ('n'=P-Y) in a single "exposed" community , what is the 95% CI for the average number of cases (μ scaled to the same amount of experience) that (would) occur in (all such) exposed communities ?	• Use tabulated CI's e.g. p 20 in this material, the CRC handbook, Documenta Geigy scientific tables, Biometrika Tables for Statisticians, ... *(Most end at c=30 or c=50)* • If have to, can use (a) trial and error on spreadsheet, or .. (b) the link between the Poisson tail areas and the tail area of the chi-square distribution.	• Same as for small numbers, or... • One of 4 approximations on p 23 (1) Wilson/Hilferty approxn. to Chi-square quantiles (X^2<-->Poisson). (2) Square-root transformation of Poisson variable. (3) 1st Principles CI from c ~ Gaussian(μ , SD = √μ) (4) (Naive) CI based on c ~ Gaussian(μ , $\hat{SD}$ = √ c). • X^2 and Likelihood Ratio (LR) methods (Miettinen Ch 10, pp 137-9)
CI for rate: $\frac{E[c]}{n}$ × N	$\frac{\textbf{CI for } \mu}{n}$ × N	$\frac{\textbf{CI for } \mu}{n}$ × N

See Liddell, FDK. Simple exact analysis of the standardized mortality ratio. Journal of Epidemiology and Community Health, March 1984, Vol 38, No. 1, pages 85-88.... on 626 website. This paper deals with SMR's but since the numerator of an SMR is treated as arising from a Poisson distribution, and the denominator as a constant, the results dealing with CI's for an SMR are also relevant just for the CI for a single Poisson parameter.

Inferences regarding a single event rate parameter: i.e. rate of events per N [$=10^x$] units of experience

data: **c "events"** counted in sample of n units of "experience"; or Binomial(c,n) if c << n. *(See again "Rothman and Walker).*

	Small no. of events	Large no. of events
Test $E[c] = E_0$ ***Example***: Is the O=2 cases of leukemia at Douglas Point statistically significantly higher than the E=0.57 cases "expected" under the null for this many person years of observation? ***Example*** What is the probability of getting 6 or more sets of twins in one school when the expected number, for schools of this size, is μ = 1.3? ***Example*** Where does the O=78 cases of cancer in the "Sour Gas" community of Alberta fall relative to E= 85.9 "expected" for "non-sour -gas" communities with the same person years of experience and at Alberta cancer rates?	**P-Value obtained by adding the individual Poisson probabilities to obtain a tail area** *(as done for Binomial and hypergeometric probabilities).* These individual probabilities are tabulated, for various 'round' values of E_0, on page 17 and in the sources listed above. E or μ = 0.57 is not tabulated but μ=0.5 and μ=0.6 are. P[2 or more events | μ=0.5] = (76+13+2)/1000 = **0.091**. P[2 or more events | μ=0.6] = (99+20+3)/1000 = **0.122**. So, P[2 or more events | μ=0.57] ≈ 0. 11 (upper tail p-value only) Instead of interpolation for non-round values of E_0, use a calculator/ spreadsheet / statistical package. Excel and SAS have Poisson probability and cumulative probability functions built in. E.g., the Excel Poisson(x, mean, cumulative) function returns a value of 0.89 when ones puts x=1, mean=0.57, cumulative = TRUE). This is the sum of the 2 tail probabilities P(0|E=0.57)= 0.57 andP(1|E=0.57) =0.32. The complement, 0.11, of the 0.89 is the upper tail p-value P(2) + P(3) + P(4) + ... So the interpolation above is quite accurate. Same procedure for c=6 vs. E=1.3 in twins data. If one sets cumulative=FALSE, the Excel function calculates the probability at the integer x only, and does not sum all of the probabilities from 0 to x. For example, setting x=9, mean=16.0 and cumulative = FALSE (or 0) yields the P(9 | μ = 16.0) = 0.21 shown in the Figure on page 18 and in row 9 of the μ=16.0 column on p 17.	- nomogram by Bailar & Ederer 1964* - 2 Gaussian approximations (from page 23) (2) **square root transformation of Poisson distribution** i.e. $z = (\sqrt{c} - \sqrt{E_0})/(0.5)$. $= (\sqrt{78} - \sqrt{85.9})/(0.5) =$ **- 0.87** (4) **asymptotic normality of c** : $z = (c - E_0) / \sqrt{E_0}$ $= (78 - 85.9) / \sqrt{85.9} =$ **- 0.85** Squaring (4) gives X^2 form (1 df) $X^2 = (c - E_0)^2 / E_0$ $= (78 - 85.9)^2 / 85.9 =$ **0.72** - Miettinen Chapter 10

* Bailar, J.C. & Ederer, F. Significance factors for the ratio of a Poisson variable to its expectation. Biometrics, Vol 20, pages 639-643, 1964.

Inference concerning comparative parameters: Rate Difference (RD) and Rate Ratio (RR)

Rate Parameters $\mathbf{R_1}$ and $\mathbf{R_0}$; Rate Difference Parameter $\mathbf{RD = R_1 - R_0}$

data: c_1 and c_0 "events" (total $c = c_1 + c_0$) in n_1 and n_0 (total=n) units of experience"; empirical rates $r_1 = \frac{c_1}{n_1}$ and $r_0 = \frac{c_0}{n_0}$;

[e.g.Rothman & Boice compare c_1=41 in n_1 =28,010 person years (PY) with c_0=15 in n_0 =19,017 person years (PY)]

	Small no. of events	Large no. of events
CI for **RD**	"Exact" methods are difficult, since t he presence of a nuisance parameter complicates matters. See papers by Suissa and by Nurminen and Miettinen. Note however that even if numerators (c_1 and c_0) are small (or even zero!) one may still have considerable precision for a rate difference: if statistical uncertainty about each rate is small, the uncertainty concerning their difference must also be small. Contrast this with situation for RR, where small numerators make RR estimates unstable. (see report by J Caro on mortality following use of low and high osmolar contrast media in radiology)	$r_1 - r_0 \pm z\sqrt{\{SE[r_1]\}^2 + \{SE[r_0]\}^2}$ in our example... $\frac{41}{28010} - \frac{15}{19017}$ $\pm 1.96\sqrt{\frac{\frac{41}{28010}(1 - \frac{41}{28010})}{28010} + \frac{\frac{15}{19017}(1 - \frac{15}{19017})}{19017}}$ Can dispense with the "1 minus small rate" term in each (binomial) variance, so the standard error of the rd simplifies to $\sqrt{\frac{c_1}{n_1^2} + \frac{c_0}{n_0^2}}$

(see Walker ; or Rothman 2002, pp 137-138)

Inference concerning comparative parameters: Rate Difference (RD) and Rate Ratio (RR)

Rate Parameters R_1 and R_2 Rate Ratio Parameter $RR = R_1 / R_0$ *See Rothman 2002, pp 137-138)*

data: c_1 and c_0 "events" (total $c = c_1 + c_0$) in n_1 and n_0 (total=n) units of experience"; empirical rates $r_1 = c_1/n_1$ & $r_0 = c_0/n_0$;

Small no. of events

CI for RR

Use distribution of c_1 conditional on $c = c_1 + c_0$ [56 in e.g. *-- not that small !*]

Conditioning on the total no. of cases, c, gets rid of one (nuisance) parameter, and lets us focus on the observed "proportion of exposed cases " (c_1/ c) and its theoretical (parameter) counterpart.

In e.g., proportion of "exposed" PY $= \frac{28010}{28010 + 19017} = 0.596 = 59.6\%$

There is a 1:1 correspondence between the expected proportion of exposed cases (call it π for short) and the RR parameter, and correspondingly between the observed proportion (p) of exposed cases and the point estimate, rr,of the rate ratio.

Under the null (RR=1), π clearly equals the proportion 0.596;

If RR > 1, this expected proportion π is higher; for example if RR=2, so that each exposed PY generates 2 times as many cases as an unexposed PY,

$$\pi = \frac{28010 \times 2}{28010 \times 2 + 19017} = 74.7\% = 0.747.$$

Thus, in our example... (and in general, $\pi = \frac{n_1 \times RR}{n_1 \times RR + n_0}$)

RR	0.25	0.50	1.00	2.00	4.00	8.00
π (proportion of exposed cases)	0.269	0.424	0.596	0.747	0.855	0.922

The observed proportion of exposed cases is p = 41/56 = 0.732; in our table, the 0.732 corresponds to an RR point estimate just below 2.

We can reverse the general formula to get RR = $\{\pi/(1-\pi)\}$ / $\{n_1/n_0\}$ = $\{\pi/(1-\pi)\}\{n_0/n_1\}$

So, in our e.g., the point estimate of RR is rr = (0.732/0.268) / (28010/19017) = 1.86 .

To obtain a CI, we treat the proportion of exposed cases, 0.732, as a binomial proportion, based on 41 "positives" out of a total of 56 cases (obviously, if the proportion were based on 8 exposed cases out of 11 cases, or 410 out of 560, the precision would be very different!)

From table/other source of CI's for proportions (see e.g. table on 607 web page), can determine that 95% CI for π is π_L=0.596 to π_U=0.842. Substitute these for the point estimate to get

RR_L = (0.596 / 0.404) / (28010/19017) = 1.00 RR_U = (0.842/0.158) / (28010/19017) = 3.61

Rothman & Walker emphasize formula $RR_{L,U} = \{\pi_{L/U} / (1-\pi_{L/U})\} / \{n_1 / n_0\}$ over basis for it.

SEE EXAMPLE IN 626 EXAM IN 2002 (0 and 41 seroconversions following vaccination vs HPV)

Large no. of events

- Use same conditional (binomial-based) formula as for small no. of events, but use Gaussian approxn. to get Binomial CI for π
- Test-based CI (Miettinen)

Uses fact that in vicinity of RR=1, can obtain SE for ln(rr) indirectly from null X^2 test statistic

X^2 statistic = square of Zstatistic = 4.33 = 2.08^2 in e.g.

Xstatistic = Zstatistic = $\frac{\ln(rr) - 0}{SE[\ln(rr)]}$

so SE[ln(rr)] = $\frac{\ln(rr)}{Xstatistic}$

CI for ln(RR) = ln(rr) ± z $\frac{\ln(rr)}{Xstatistic}$

CI for RR: rr to power $[1 \pm \frac{z}{Xstatistic}]$

= 1.86 to power of [1 ± 1.96/2.08]
= 1.04 to 3.32 in e.g.

- Var [ln(rr)] = $\frac{1}{c_1} + \frac{1}{c_0} + \frac{1}{\infty} + \frac{1}{\infty}$ (Woolf)

CI for **RR** = rr exp[$\pm z \sqrt{\frac{1}{c_1} + \frac{1}{c_0}}$]

1.96 $(1/41+1/15)^{1/2}$ =0.59 in e.g. ;

so exp[0.59]=1.81; So CI for RR

=.86 / 1.81 to 1.86*1.81 = (1.02,3.35)

Precision for ln(RR) estimate depends on numbers of events c_1 and c_0.

Inference concerning comparative parameters: Rate Difference (RD) and Rate Ratio (RR)

Rate Parameters $\mathbf{R_1}$ and $\mathbf{R_2}$ Rate Difference Parameter $\mathbf{RD = R_1 - R_0}$ Rate Ratio Parameter $\mathbf{RR = R_1 / R_0}$

data: c_1 and c_0 "events" (total $c = c_1 + c_0$) in n_1 and n_0 (total=n) units of experience"; empirical rates $r_1 = c_1/n_1$ & $r_0 = c_0/n_0$;

	Small no. of events	Large no. of events
test of RD=0 or RR=1	• Null distribution of c_1 conditional on c $c_1 \mid c \sim$ Binomial, with c "trials", (see above) each with null probability $\pi = \frac{\mathbf{RR} \times n_1}{\mathbf{RR} \times n_1 + n_0}$. e.g. If RR =1 (RD=0) would expect the 56 cases to split into "exposed" and "unexposed" in the proportions 27010/(27010+19017) = 0.596 and 1-0.596=0.404 respectively. Can test if the observed proportion 41/56 = 0.732 is significantly different from this null expectation using a Binomial distribution with "n"=56 and π=0.596. Can use the Excel Binomial function with x=40,mean=0.596,cumulative=TRUE, to get the sum of all the probabilities up to and including 40. Subtract this quantity 0.976 from 1 to get the probability 0.024 of 41 or more (upper tail area). Double this for a 2-sided test. • Unconditional test for proportions / rates (Suissa)	• Use same "c_1 conditional on c" test but use Gaussian approx^n to Binomial (c, π) e.g. $z = \frac{[41/56 =]0.732 - 0.596}{\sqrt{0.596 \times 0.404/56}} = 2.08$ P(Z > z) = 0.019 (upper tail area). Double for 2-sided test. • $z = \frac{[r_1 - r_2] - RD_0}{\sqrt{\{SE[r_1 \mid H_0\}^2 + \{SE[r_2 \mid H_0\}^2}}$ {*SE's use r = Σc/Σn [pooled data]} • $X^2 = \frac{\{c_1 - E[c_1 \mid H_0]\}^2}{E[c_1 \mid H_0]} + \frac{\{c_0 - E[c_0 \mid H_0]\}^2}{E[c_0 \mid H_0]}$ $= \frac{\{c_1 - E[c_1 \mid H_0]\}^2}{Var[c_1 \mid H_0]}$ (Mantel-Haenszel version) *See my notes on Chi-square tests in on chapter 8 in 607 course*