

Parameters: π_1 and π_2 's are proportions / prevalences / risks

3 Comparative measures / parameters:

	COMPARATIVE PARAMETER	estimate	New Scale
(Risk or Prevalence) Difference	$\pi_1 - \pi_2$	$p_1 - p_2$	
			cf. Rothman2002 p 135 Eqn 7-2 <<<<<<

(Risk or Prevalence) Ratio	$\frac{\pi_1}{\pi_2}$	$\frac{p_1}{p_2}$	$\log\left[\frac{p_1}{p_2}\right] = \log[p_1] - \log[p_2]$
			cf. Rothman2002 p 135 Eqn 7-3 <<<<<<

Odds Ratio	$\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$	$\frac{odds_1}{odds_2}$	$\log\left[\frac{odds_1}{odds_2}\right]$
			$= \log[odds_1] - \log[odds_2]$
			$= \text{logit}_1 - \text{logit}_2$
			cf. Rothman2002 p 139 Eqn 7-6 <<<<<<

Large-sample CI for COMPARATIVE MEASURE / PARAMETER (if 2 estimates are uncorrelated)

IN GENERAL (if calculations in transformed scale, must back-transform)
 $\text{estimate}_1 - \text{estimate}_2 \pm z \text{SE}[\text{estimate}_1 - \text{estimate}_2]$
 $\text{estimate}_1 - \text{estimate}_2 \pm z \text{Sqrt}[\text{Var}[\text{estimate}_1] + \text{Var}[\text{estimate}_2]]$

IN PARTICULAR
 $p_1 - p_2 \pm z \text{SE}[p_1 - p_2]$ *Remember: SE's don't add !*
 $= p_1 - p_2 \pm z \sqrt{\text{SE}^2[p_1] + \text{SE}^2[p_2]}$ *Their squares do !*
 $= p_1 - p_2 \pm z \sqrt{\frac{p_1[1-p_1]}{n_1} + \frac{p_2[1-p_2]}{n_2}}$

$\text{anti-log}[\log[p_1/p_2] \pm z \text{SE}[\log[p_1] - \log[p_2]]]$
 $= \text{anti-log}[\log[p_1/p_2] \pm z \sqrt{\text{SE}^2[\log[p_1]] + \text{SE}^2[\log[p_2]]}]$
 From 8.1: $\text{SE}^2[\log[p_1]] = \text{Var}[\log[p_1]] = 1/\#\text{positive}_1 - 1/\#\text{total}_1$
 $\text{SE}^2[\log[p_2]] = \text{Var}[\log[p_2]] = 1/\#\text{positive}_2 - 1/\#\text{total}_2$

$\text{anti-log}[\log[\text{oddsRatio}] \pm z \text{SE}[\text{logit}_1 - \text{logit}_2]]$
 $= \text{anti-log}[\log[\text{oddsRatio}] \pm z \sqrt{\text{SE}^2[\text{logit}_1] + \text{SE}^2[\text{logit}_2]}]$
 From 8.1: $\text{SE}^2[\text{logit}_1] = \text{Var}[\text{logit}_1] = 1/\#\text{positive}_1 + 1/\#\text{negative}_1$
 $\text{SE}^2[\text{logit}_2] = \text{Var}[\text{logit}_2] = 1/\#\text{positive}_2 + 1/\#\text{negative}_2$
 $\text{Var}[\log \text{ of OR est.}] = 1/a + 1/b + 1/c + 1/d \implies \text{"Woolf's Method"}: \text{CI}[\text{ODDSRATIO}]$

Large-sample

Test of $p_1 = p_2$ equivalent to Test of $p_1 - p_2 = 0$
 (Risk or Prevalence Difference = 0)

same as Test of $\frac{p_1}{n_1} = \frac{p_2}{n_2}$
 (Risk or Prevalence Ratio = 1)

same as Test of $\frac{p_1 / \{1 - p_1\}}{p_2 / \{1 - p_2\}} = 1$
 (ODDS Ratio = 1)

$$z = \frac{p_1 - p_2 - \{=0\}}{SE[p_1 - p_2]} \quad \dagger$$

$$= \frac{p_1 - p_2}{\sqrt{\frac{p[1-p]}{n_1} + \frac{p[1-p]}{n_2}}} \quad \dagger \dagger$$

$$= \frac{p_1 - p_2}{\sqrt{p[1-p] \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}}} = \frac{p_1 - p_2}{p[1-p] \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

[use estimate $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{\text{total +ve}}{\text{total}}$ in test]

† Continuity correction: use $|p_1 - p_2| - [1/(2n_1) + 1/(2n_2)]$ in numerator

† † Variances add if the proportions are uncorrelated

Examples:

0 The generic 2x2 contingency table:

	+ve	-ve	both
sample 1	y_1 (%)	$n_1 - y_1$	n_1 (100%)
sample 2	y_2 (%)	$n_2 - y_2$	n_2 (100%)
total	y (%)	$n - y$	n (100%)

1 Bromocriptine for unexplained 1° infertility (BMJ 1979)

	became pregnant	did not	total no. couples
Bromocriptine	7 (29%)	17	24 (100%)
Placebo	5 (22%)	18	23 (100%)
total	12 (26%)	35	47 (100%)

2 Vitamin C and the common cold (CMAJ Sept 1972 p 503)

	no colds	1 cold	total subjects
Vitamin C	105 (26%)	302	407 (100%)
Placebo	76 (18%)	335	411 (100%)
total	181 (22%)	637	818 (100%)

3 Stroke Unit vs Medical Unit for Acute Stroke in elderly? Patient status at hospital discharge (BMJ 27 Sept 1980)

	indept.	dependent	total no. pts
Stroke Unit	67 (66%)	34	101 (100%)
Medical Unit	46 (51%)	45	91 (100%)
total	113 (59%)	79	192 (100%)

Stroke Unit vs Medical Unit

$$\begin{aligned} \text{95\%CI on } \underline{\quad} : & 0.66 - 0.51 \pm z \sqrt{\frac{0.66 \cdot 0.34}{101} + \frac{0.51 \cdot 0.49}{91}} \\ & = 0.15 \pm 1.96 \times 0.07 \\ & = 0.15 \pm 0.14 \end{aligned}$$

test $=0$: [carrying several decimal places, for comparison with 2 later]

$$z = \frac{0.6634 - 0.5054}{\sqrt{0.5885 \cdot 0.4115 \cdot \left\{ \frac{1}{101} + \frac{1}{91} \right\}}} = \frac{0.1580}{0.0711} = 2.22$$

[estimate of hypothesized common $\pi = \frac{\text{total +ve}}{\text{total}} = \frac{113}{192} = 0.5885$]

P = Prob[|Z| 2.22] = 0.026 (2-sided)

Vitamin C and the common cold

$$\begin{aligned} \text{95\%CI on } \underline{\quad} : & 0.26 - 0.18 \pm z \sqrt{\frac{0.26 \cdot 0.74}{407} + \frac{0.18 \cdot 0.81}{411}} \\ & = 0.08 \pm 1.96 \times 0.03 \\ & = 0.08 \pm 0.06 \end{aligned}$$

test $=0$: [carrying several decimal places, for comparison with 2 later]

$$z = \frac{0.258 - 0.185}{\sqrt{0.221 [0.779] \left\{ \frac{1}{407} + \frac{1}{411} \right\}}} = \frac{0.073}{0.029} = 2.52 \text{ (2.517 = 6.337)}$$

P=0.006 (1-sided); P=0.012 (1-sided)

Rx for primary infertility

$$\begin{aligned} \text{95\%CI on } \underline{\quad} : & 0.29 - 0.22 \pm z \sqrt{\frac{0.29 \cdot 0.71}{24} + \frac{0.22 \cdot 0.78}{23}} \\ & = 0.07 \pm 1.96 \times 0.13 \\ & = 0.07 \pm 0.25 \end{aligned}$$

test $=0$:

$$z = \frac{0.29 - 0.22}{\sqrt{0.26 [0.74] \left\{ \frac{1}{24} + \frac{1}{23} \right\}}} = \frac{0.07}{0.13} = 0.55$$

P=0.58 (2-sided)

Fall in BP with Reduction in Dietary Salt

Response of interest: Y = Achieve DBP < 90 ? [0 / 1]

H_0 : (Y=1 | Normal Sodium Diet) = (Y=1 | Low Sodium Diet)

H_{alt} : (Y=1 | Normal Sodium Diet) (Y=1 | Low Sodium Diet)

=0.05 (2-sided);

Proportion achieving DBP < 90 mm

"Normal" Group	"Low" Group	
11/50	17/53	
(22%)	(32%)	$z = \frac{0.32 - 0.22}{\sqrt{0.27 \cdot 0.73 \cdot \left[\frac{1}{53} + \frac{1}{50} \right]}}$

i.e. |z| = 1.14 which is < Z = 1.96 and so observed difference of 10% is "N.S."

CI for Risk Ratio (Rel. RISK) or Prev. Ratio cf. Rothman2002 p135

Example: Vitamin C and the common cold (CMAJ Sept 1972 p 503) ... REVISITED

	no colds	1 cold	total subjects
Vitamin C	105 (26%)	302(74%)	407 (100%)
Placebo	76 (18%)	335(82%)	411 (100%)
total	181 (22%)	637(78%)	818 (100%)

$$\hat{RR} = \frac{\text{Prob}[\text{cold} | \text{Vitamin C}]}{\text{Prob}[\text{cold} | \text{Placebo}]} = \frac{74\%}{82\%} = 0.91$$

$$\text{CI}[\hat{RR}]: \text{antilog}\{ \log[0.91] \pm z \text{SE}[\log[p_1] - \log[p_2]] \}$$

$$= \text{antilog}\{ \log[0.91] \pm z \sqrt{\text{SE}^2[\log[p_1]] + \text{SE}^2[\log[p_2]]} \}$$

From 8.1: $\text{SE}^2[\log[p_1]] = \text{Var}[\log[p_1]] = 1/302 - 1/407 = 0.000854$
 $\text{SE}^2[\log[p_2]] = \text{Var}[\log[p_2]] = 1/335 - 1/411 = 0.000552$

$$\text{CI}[\hat{RR}]: \text{antilog}\{ \log[0.91] \pm z \text{Sqrt}[0.000854 + 0.000552] \}$$

$$= \text{antilog}\{ \log[0.91] \pm 0.073 \} = 0.85 \text{ to } 0.98$$

Shortcut: calculate $\exp\{ z \times \text{SE}[\log \hat{RR}] \}$ and use it as a multiplier and divider of \hat{RR} . In our e.g., $\exp\{ z \times \text{SE}[\log \hat{RR}] \} = \exp\{0.073\} = 1.076$.

Thus $\{RR_{\text{LOWER}}, RR_{\text{UPPER}}\} = \{0.91 / 1.076, 0.91 \times 1.076\} = \{0.85, 0.98\}$
 You can use this shortcut whenever you are working with log-based CI's that you convert back to the original scale, there they become "multiply-divide" symmetric rather than "plus-minus" symmetric.

<p>From SAS</p> <pre>PROC FORMAT; VALUE onefirst 0="z_0" 1="a_1"; DATA CI_RR_OR; INPUT vitC cold n_people; LINES; 1 1 302 1 0 105 0 1 335 0 0 76 ; PROC FREQ data=CI_RR_OR ORDER=FORMATTED; TABLES vitC*cold / CMH; WEIGHT n_people; FORMAT vitC cold onefirst. ;RUN;</pre>	<p>From Stata</p> <pre>Immediate: csi 302 335 105 76 cs stands for 'cohort study' input vit_c cold n_people 1 1 302 1 0 105 0 1 335 0 0 76 end cs cold vit_c [freq = n_people]</pre>
--	--

CI for ODDS RATIO cf. Rothman2002 p139

	Vitamin C	Placebo:
had cold(s)	302	335
avoided colds	105	76
# with cold(s) for every 1 who avoided colds	2.88 (:1)	4.41 (:1)
odds of cold(s)	2.88	4.41

$$\text{odds Ratio} = 2.88 / 4.41 = 0.65 \gg \hat{OR} = 0.65$$

$$\text{CI}[\hat{OR}] = \text{anti-log} [\log[\text{oddsRatio}] \pm z \text{SE}[\text{logit}_1 - \text{logit}_2]]$$

From 8.1: $\text{SE}^2[\text{logit}_1] = 1/\#\text{positive}_1 + 1/\#\text{negative}_1$

$$\text{SE}^2[\text{logit}_2] = 1/\#\text{positive}_2 + 1/\#\text{negative}_2$$

$$\text{SE}[\text{logit}_1 - \text{logit}_2] = \text{Sqrt}[(1/302 + 1/105) + (1/335 + 1/76)] = 0.17$$

$$z \text{SE}[\text{logit}_1 - \text{logit}_2] = 1.96 \times 0.17 = 0.33$$

$$\text{anti-log} [\log[0.65] \pm 0.33] = \exp[-0.43 \pm 0.33] = 0.47 \text{ to } 0.90$$

$$\gg \gg \gg \hat{OR} = 0.65; \text{CI}[\hat{OR}] = \{0.47 \text{ to } 0.90\}$$

<p>From SAS</p> <pre><<<< See statements for RR (output gives both RR and OR) Be CAREFUL as to rows/cols Index exposure category must be 1st row; reference exposure category must be 2nd If necessary, use FORMAT to have table come out this way... (note trick to reverse rows/cols) SAS doesn't know if cc or cohort</pre>	<p>From Stata</p> <pre>Immediate: cci 302 335 105 76,woolf cc stands for 'case control study' input vit_c cold n_people 1 1 302 1 0 105 0 1 335 0 0 76 end cc cold vit_c [freq =n_people],woolf</pre>
--	---

"Test-based" CI's ... IN GENERAL

Preamble

In 1959, when Mantel and Haenszel developed their summary Odds Ratio measure over 2 or more strata, they did not supply a CI to accompany this point estimate. From 1955 onwards, the main competitor was the weighted average (in the log OR scale) and accompanying CI obtained by Woolf. But this latter method has problems with strata where one or more cell frequencies are zero. In 1976, Miettinen developed the "test-based" method for epidemiologic situations where the summary point estimate is easily calculated, the standard error estimate is unknown or hard to compute, but where a statistical test of the null value of the parameter of interest (derived by aggregating a "sub-statistic" from each stratum) is already available. Although the 1886 development, by Robins, Breslow and Greenland, of a direct standard error for the log of the Mantel-Haenszel OR estimator, the "test-based" CI is still used (see A&B KKM).

Even though its main usefulness is for summaries *over strata*, the idea can be explained using a simpler and familiar (single stratum) example, the comparison of two independent means using a z-test with large df (the principle does not depend on t vs. z). Suppose all that was reported was the difference in sample means, and the 2-sided p-value associated with a test of the null hypothesis that the mean difference was zero. From the sample means, and the p-value, how could we obtain a 95%CI for the difference in the "population" means? The trick is to

- 1 work back (using a table of the normal distribution) from the p-value to the corresponding value of the z-statistic (the number of standard errors that the difference in sample means is from zero);
- 2 divide this observed difference by the observed z value, to get the standard error of the difference in sample means, and
- 3 use the observed difference, and the desired multiple (1.645 for 90% CI, 1.96 for 95% etc.) to create the CI.

The same procedure is directly applicable for the difference of two independently estimated proportions. If one tests the (null) difference using a z-test, one can obtain the SE of the difference by dividing the observed difference in proportions by the z statistic; if the difference was tested by a chi-square statistic, one can obtain the z-statistic by taking the square root of the observed chi-square value (authors call this square root an observed 'chi' value). Either way, the observed z-value leads directly to the SE, and from there to the CI.

This is worked out in the next example, where it is assumed that the null hypothesis is tested via a chi-squared (χ^2) test

"Test-based" CI's for... IN PARTICULAR

- **Difference** of 2 proportions $p_1 - p_2$ (**Risk** or Prevalence **Difference**)

Observe : p_1 & p_2 and (maybe via p-value) the calculated value of χ^2

This implies that

$$\text{Sqrt}[\text{observed } \chi^2 \text{ value}] = \text{observed } x \text{ value} = \text{observed } z \text{ value};$$

But... observed z statistic = $(p_1 - p_2) / \text{SE} [p_1 - p_2]$.

$$\text{So... } \text{SE} [p_1 - p_2] = \frac{p_1 - p_2}{\text{observed } z \text{ statistic}} \quad \text{use +ve sign}$$

95% CI for $p_1 - p_2$:

$$(p_1 - p_2) \pm \{z \text{ value for 95\%}\} \times \text{SE}[p_1 - p_2]$$

i.e.,...

$$(p_1 - p_2) \pm \{z \text{ value for 95\%}\} \times \frac{p_1 - p_2}{\text{observed } z \text{ statistic}}$$

i.e., *after re-arranging terms..*

$$(p_1 - p_2) \left\{ 1 \pm \frac{z \text{ value for 95\%}}{\text{observed } z \text{ statistic}} \right\}$$

or, in terms of a reported chi-square statistic

$$(p_1 - p_2) \left\{ 1 \pm \frac{z \text{ value for 95\%}}{\text{Sqrt}[\text{observed } \chi\text{-square statistic}]} \right\}$$

See Section 12.3 of Miettinen's "Theoretical Epidemiology"

Technically, when the variance is a function of the parameter (as is the case with binary response data), the test-based CI is most accurate close to the Null. However, as you can verify by comparing test-based CIs with CI's derived in other ways, the inaccuracies are not as extreme as textbooks and manuals (e.g. Stata) suggest.

"Test-based" CI's for... IN PARTICULAR

- **Ratio** of 2 proportions p_1/p_2

(**Risk Ratio; Prevalence Ratio; Relative Risk; "RR"**)

Observe :(i) $rr = p_1/p_2$ and

(ii) (maybe via p-value) the value of χ^2 statistic ($H_0: RR=1$)

>>> $\text{Sqrt}[\text{observed } \chi^2 \text{ value}] = \text{observed } x \text{ value} = \text{observed } z \text{ value}$

In log scale, in relation to $\log[RR_{null}] = 0$, observed z value would be:

$$\text{observed } z \text{ value} = (\log[rr] - 0) / \text{SE}[\log[rr]]$$

This implies that

$$\text{SE}[\log[rr]] = \log[rr] / \text{observed } z \text{ value} \quad \textit{use +ve sign}$$

95% CI for **log[RR]**:

$$\log[rr] \pm \{z \text{ value for } 95\% \} \times \text{SE}[\log[rr]]$$

i.e.,...

$$\log[rr] \pm \{z \text{ value for } 95\% \} \times \frac{\log[rr]}{\text{observed } z \text{ value}}$$

i.e., *after re-arranging terms..*

$$\log[rr] \times \left\{ 1 \pm \frac{z \text{ value for } 95\%}{\text{observed } z \text{ statistic}} \right\}$$

Going back to RR scale, by taking antilogs*...

95% CI for **RR**:

$$rr \text{ to power of } \left\{ 1 \pm \frac{z \text{ value for } 95\%}{\text{observed } z \text{ statistic}} \right\}$$

See Section13.3 of Miettinen's "Theoretical Epidemiology"

* $\text{antilog}[\log[a] \times b] = \exp[\log[a] \times b] = \{ \exp[\log[a]] \}$ to power of $b = a$ to power of b

"Test-based" CI's for... IN PARTICULAR

- **Ratio** of 2 odds $p_1/[1-p_1]$ and $p_2/[1-p_2]$

(**Odds Ratio; "OR"**)

Observe :(i) $or = p_1/[1-p_1] / p_2/[1-p_2]$ (" $a \times d / b \times c$ ") and

(ii) (maybe via p-value) the value of χ^2 statistic ($H_0: OR=1$)

>>> $\text{Sqrt}[\text{observed } \chi^2 \text{ value}] = \text{observed } x \text{ value} = \text{observed } z \text{ value}$

In log scale, in relation to $\log[OR_{null}] = 0$, observed z value would be:

$$\text{observed } z \text{ value} = (\log[or] - 0) / \text{SE}[\log[or]]$$

This implies that

$$\text{SE}[\log[or]] = \log[or] / \text{observed } z \text{ value} \quad \textit{use +ve sign}$$

95% CI for **log[OR]**:

$$\log[or] \pm \{z \text{ value for } 95\% \} \times \text{SE}[\log[or]]$$

i.e.,...

$$\log[or] \pm \{z \text{ value for } 95\% \} \times \frac{\log[or]}{\text{observed } z \text{ value}}$$

i.e., *after re-arranging terms..*

$$\log[or] \times \left\{ 1 \pm \frac{z \text{ value for } 95\%}{\text{observed } z \text{ statistic}} \right\}$$

Going back to OR scale, by taking antilogs*...

95% CI for **OR**:

$$or \text{ to power of } \left\{ 1 \pm \frac{z \text{ value for } 95\%}{\text{observed } z \text{ statistic}} \right\}$$

See Section13.3 of Miettinen's "Theoretical Epidemiology"

Sample Size considerations... CI(1 - 2)

n's to produce CI for difference in π 's of pre specified margin of error m at stated confidence level

• large-sample CI: $p_1 - p_2 \pm Z SE(p_1 - p_2) = p_1 - p_2 \pm m$

• $SE(p_1 - p_2) = \sqrt{\frac{p_1\{1-p_1\}}{n_1} + \frac{p_2\{1-p_2\}}{n_2}}$

Simplify by using an average p;

if use equal n's, then

$$n \text{ per group} = \frac{2p\{1-p\} Z_{/2}^2}{[\text{margin of error}]^2}$$

M&M use the fact that if $p = 1/2$ then $p(1-p) = 1/4$, and so $2p(1-p) = 1/2$, so the above equation becomes

$$[\text{max}] n \text{ per group} = \frac{Z_{/2}^2}{2 \times [\text{margin of error}]^2}$$

Sample Size considerations... Test involving T and C

Test $H_0: \pi_T = \pi_C$ vs $H_a: \pi_T \neq \pi_C$:

n's for power $1-\beta$ if $\pi_T = \pi_C + \Delta$; $\text{prob}[\text{type I error}] = \alpha$

n per group

$$= \frac{\{Z_{/2} \sqrt{2\pi_C\{1-\pi_C\}} - Z_{\beta} \sqrt{\pi_C\{1-\pi_C\} + \pi_T\{1-\pi_T\}}\}^2}{\Delta^2}$$

(See Colton p 168)

$$2(Z_{/2} - Z_{\beta})^2 \{ \sqrt{\pi_C\{1-\pi_C\}} \}^2$$

$$= 2\{Z_{/2} - Z_{\beta}\}^2 \{ \frac{0.1}{\Delta} \}^2$$

e.g.

$\alpha=0.05$ (2-sided) & $\beta=0.2$... $Z_{/2} = 1.96$; $Z_{\beta} = -0.84$,

$$2(Z_{/2} - Z_{\beta})^2 = 2\{1.96 - (-0.84)\}^2 = 16, \text{ i.e.}$$

$$n \text{ per group} = 16 \cdot \pi_C\{1-\pi_C\} / \Delta^2$$

**So $n \approx 100$ for T group and $n \approx 100$ for C group
if $\pi_T = 0.6$ and $\pi_C = 0.4$**

See Sample Size Requirements for Comparison of 2 Proportions (from text by Smith and Morrow) under Resources for Chapter 8.

Effect of Unequal Sample Sizes ($n_1 \neq n_2$) on precision of estimated differences

If we write the SE of an estimated difference in mean responses as $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, where σ is the (average) per unit variability of the response, then we can establish the following principles:

1 If costs and other factors (including unit variability) are equal, and if both types of units are equally scarce or equally plentiful, then for a given total sample size of $n = n_1 + n_2$, an equal division of n i.e. $n_1 = n_2$ is preferable since it yields a smaller SE(estimated difference in means) than any non-symmetric division. However, the SE is relatively unaffected until the ratio exceeds 70:30. This is seen in the following table which gives the value of $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ = SE(estimated difference in means) for various combinations of n_1 and n_2 adding to 100 (the 100 itself is arbitrary) and assuming $\sigma = 1$ (also arbitrary).

2 If one type of unit is much scarcer, and thus the limiting factor, then it makes sense to choose all (say n_1) of the available scarcer units, and some $n_2 \ll n_1$ of the other type. The greater is n_2 , the smaller the SE of the estimated difference. However, there is a 'law of diminishing returns' once n_2 is more than a few multiples of n_1 . This is seen in the following table which gives the value of $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ for n_1 fixed (arbitrarily) at 100 and n_2 ranging from 1 x n_1 to 100 x n_1 ; again, we assume $\sigma = 1$.

n_1	n_2	SE(estimated difference in means)	%Increase in SE over SE(50:50) *
50	50	0.200	-----
60	40	0.204	2.1%
65	35	0.210	4.8%
70	30	0.218	9.1%
75	25	0.231	15.5%
80	20	0.250	25.0%
85	15	0.280	40.0%

n_1	n_2	Ratio (K)	SE($\hat{\mu}_1 - \hat{\mu}_2$)	SE _{K:1} as % of SE(1:1)	SE _{K:1} as % of SE(:1) *
50	50	1.0	0.2000	-	1.414
50	75	1.5	0.1825	91.3%	1.290
50	100	2.0	0.1732	86.6%	1.225
50	150	3.0	0.1633	81.6%	1.155
50	200	4.0	0.1581	79.1%	1.118
50	250	5.0	0.1549	77.5%	1.095
50	300	6.0	0.1527	76.4%	1.080
50	400	8.0	0.1500	75.0%	1.061
50	500	10.0	0.1483	74.2%	1.049
50	1000	20.0	0.1449	72.4%	1.025
50	5000	100.0	0.1421	71.1%	1.005
50			0.1414	70.7%	1

* if sample sizes are $n_1:(1-n_1)$, the % increase is $50 / \sqrt{1-n_1}$.

* calculated as $\sqrt{\frac{K+1}{K}}$; 'efficiency' = $\sqrt{\frac{K}{K+1}}$

Note: these principles apply to both measurement and count data

Sample size calculation when using unequal sample sizes to estimate / test difference in 2 means or proportions

For power (sensitivity) $1 - \beta$, and specificity $1 - \alpha$ (2-sided), the sample sizes n_1 and n_2 have to be such that

$$Z_{1-\beta/2} SE(\bar{x}_1 - \bar{x}_2) - Z_{1-\alpha/2} SE(\bar{x}_1 - \bar{x}_2) = \delta$$

(if $\delta < 0.5$, then $Z_{1-\beta/2}$ will be negative). If we assume equal per unit variability, σ , of the x's in the 2 populations, we can write the requirement as

$$Z_{1-\beta/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} - Z_{1-\alpha/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{\delta}{\sigma}$$

If we rewrite $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ as $\sqrt{\frac{1}{n_1} \left\{ 1 + \frac{n_1}{n_2} \right\}}$

and rearrange the inequality, we get

$$n_1 = \left\{ 1 + \frac{n_1}{n_2} \right\} (Z_{1-\beta/2} - Z_{1-\alpha/2})^2 \left\{ \frac{\delta}{\sigma} \right\}^2$$

or, denoting $\frac{n_2}{n_1}$ by K ,

$$n_1 = \left\{ 1 + \frac{1}{K} \right\} (Z_{1-\beta/2} - Z_{1-\alpha/2})^2 \left\{ \frac{\delta}{\sigma} \right\}^2$$

i.e.

$$n_1 = \left\{ \frac{K+1}{K} \right\} (Z_{1-\beta/2} - Z_{1-\alpha/2})^2 \left\{ \frac{\delta}{\sigma} \right\}^2$$

Notes:

a. If $K=1$, so that $n_1=n_2$, then we get the familiar "2" at the front of the sample size formula.

b. The same factor applies for **proportions**:

$$\text{If we use } \sigma_{0/1} = \sqrt{\bar{\pi} [1 - \bar{\pi}]}$$

as an "average" standard deviation for the individual 0's and 1's in each population, i.e.

$$\sigma_{0/1} = \sqrt{\bar{\pi} [1 - \bar{\pi}]}$$

then, as we get the approximate formula:

$$n_1 = \left\{ \frac{K+1}{K} \right\} (Z_{1-\beta/2} - Z_{1-\alpha/2})^2 \left\{ \frac{\bar{\pi} [1 - \bar{\pi}]}{2} \right\}$$

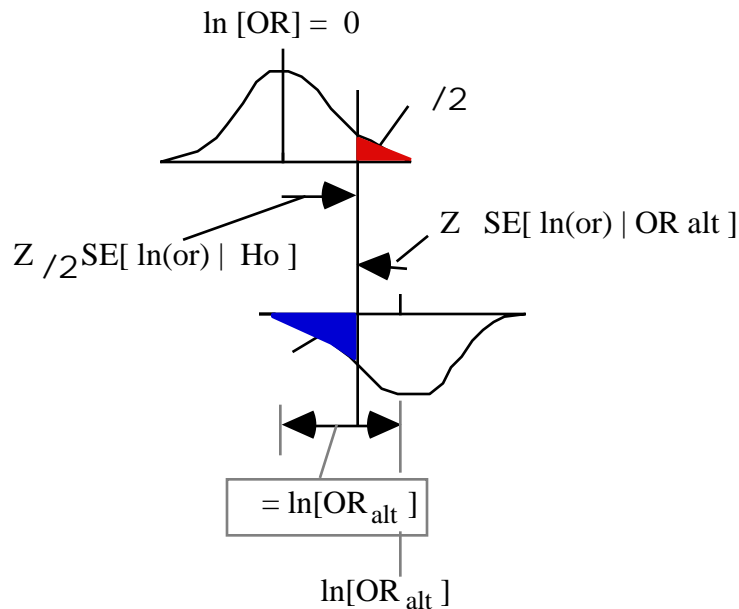
Sample Size considerations... Test involving OR
Test $H_0: OR = 1$ vs. $H_a: OR \neq OR$:
n's for power $1-\beta$ if $OR = OR_{alt}$; $prob[type\ I\ error] = \alpha$

Here I use ln for natural log (elsewhere I have used \log ; I use them interchangeably)

Work in ln (or) scale; $SE[ln(or)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$

Need $Z_{/2} SE[ln(or)]_0 + Z SE[ln(or)]_{alt} < "$ "

where " " = $ln(OR_{alt})$



Substitute expected a, b, c, d values under null and alt. into SE's and solve for numbers of cases and controls.

References: Schlesselman, Breslow and Day, Volume II, ...

Key points

$ln[or]$ most precise when all 4 cells are of equal size; so...

- 1 increasing the control:case ratio leads to diminishing marginal gains in precision.

To see this... examine the function

$$\frac{1}{\# \text{ of cases}} + \frac{1}{\text{multiple of this } \# \text{ of controls}}$$

for various values of "multiple"

[like we did back in Chapter 8, for "effect of unequal sample sizes"]

- 2 The more unequal the distribution of the etiologic / preventive factor, the less precise the estimate

Examine the functions

$$\frac{1}{\# \text{ of exposed cases}} + \frac{1}{\# \text{ of unexposed cases}}$$

and

$$\frac{1}{\# \text{ of exposed controls}} + \frac{1}{\# \text{ of unexposed controls}}$$

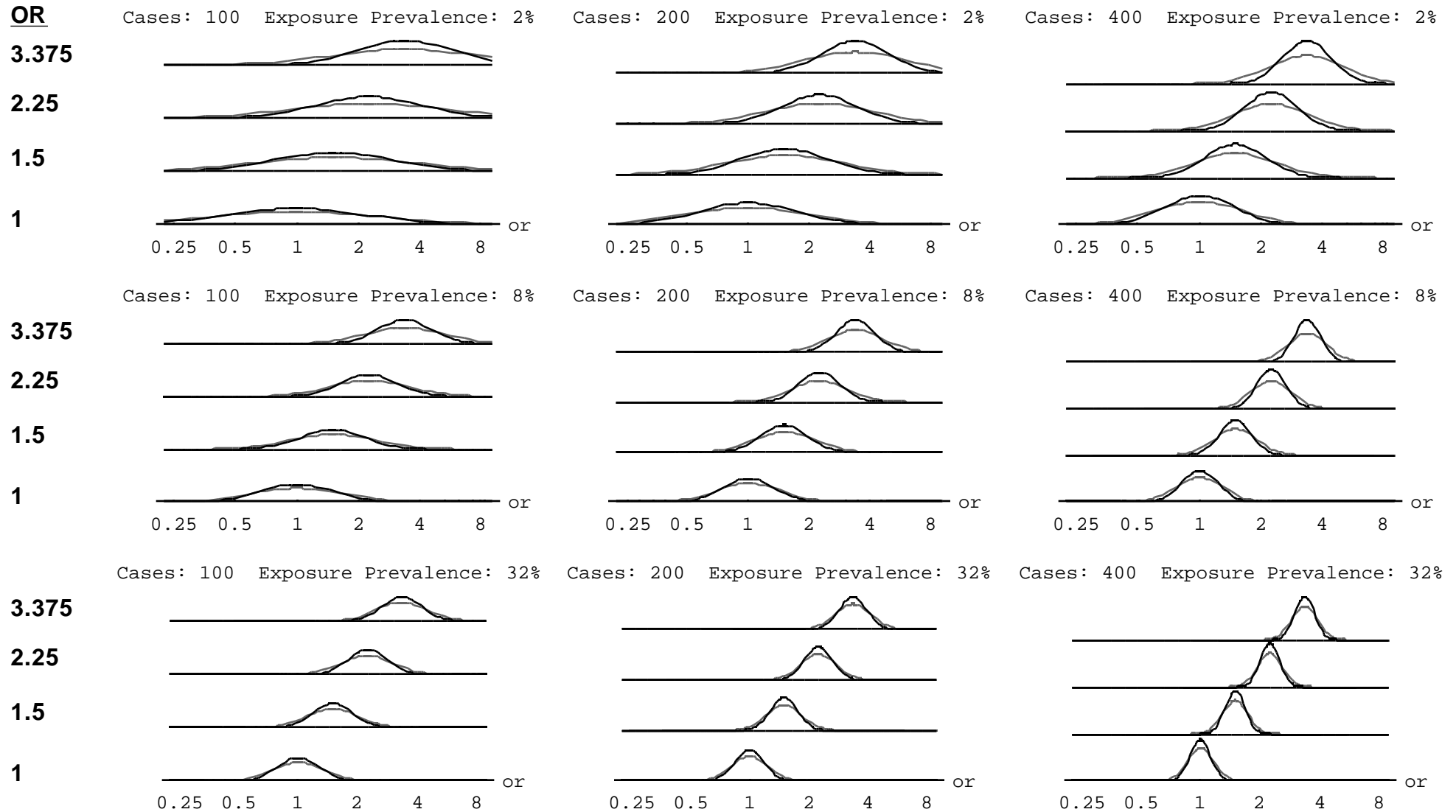
Reading graphs on next page (Note log scale for observed or)

Take as an example the study in the middle panel, with 200 cases, and an exposure prevalence of 8%. Say that the Type I error rate is set at $\alpha=0.05$ (2sided) so that the upper critical value (the one that cuts off the top 2.5% of the null distribution) is close to $or = 2$. Draw a vertical line at this critical value, and examine how much of each non-null distribution falls to the right of this critical value. This area to the right of the critical value is the power of the study, i.e., the probability of obtaining a significant or, when in fact the indicated non-null value of OR is correct. Two curves at each OR value are for studies with 1(grey) and 4(black) controls/case. Note that OR values 1, 1.5, 2.25 and 3.375 are also on a log scale.

Power larger if...

- i non-null $OR \gg 1$ (cf 2.5 vs 2.25 vs 3.375)
- ii exposure common (cf 2% vs 8% vs 32%) and not near universal)
- iii use more cases (cf 100 vs 200 vs 400), and controls/case (1 vs 4)

Factors affecting variability of estimates from, and statistical power of, case-control studies



The "Exact" Test for 2 x 2 tables [material taken from A&B §4.9]

Even with the continuity correction there will be some doubt about the adequacy of the χ^2 approximation when the frequencies are particularly small. An exact test was suggested almost simultaneously in the mid-1930s by R. A. Fisher, J. O. Irwin and F. Yates. It consists in calculating the exact probabilities of the possible tables described in the previous subsection. The probability of a table with frequencies

a	b	r ₁
c	d	r ₂
c ₁	c ₂	N

is given by the formula

$$P[a | r_1, r_2, c_1, c_2] = \frac{r_1! r_2! c_1! c_2!}{N! a! b! c! d!} \quad (4.25)$$

This is, in fact, the probability of the observed cell frequencies *conditional* on the observed marginal totals, under the null hypothesis of no association between the row and column classifications. Given any observed table, the probabilities of all tables with the same marginal totals can be calculated, and the P value for the significance test calculated by summation. Example 4.14 illustrates the calculations and some of the difficulties of interpretation which may arise. The data in Table 4.6, due to M. Hellman, are discussed by Yates (1934).

Table 4.6 Data on malocclusion of teeth in infants (Yates, 1934)

	Infants with		
	Normal teeth	Malocclusion	Total
Breast-fed	4	16	20
Bottle-ed	1	21	22
Total	5	37	42

There are six possible tables with the same marginal totals as those observed. since neither a nor c (in the notation given above) can fall below 0 or exceed 5, the smallest marginal total in the table. The cell frequencies in each of these tables are shown in Table 4.7. Below them are shown the probabilities of these tables, calculated under the null hypothesis.

Table 4.7 Cell frequencies in tables with the same marginal totals as those in Table 4.6

	0	1	2	3	4	5
a	20	19	18	17	16	15
c	17	18	19	20	21	22
c ₁	37	37	37	37	37	37
P _a	0.0310	0.1720	0.3440	0.3096	0.1253	0.0182

The Probabilities of the various tables are calculated in the following way*: the probability that a = 0 is, from (4.25),

$$P_0 = \frac{20! 22! 5! 37!}{42! 0! 20! 5! 7!} = 0.03096.$$

Tables of log factorials (Fisher and Yates, 1963, Table XXX) are often useful for this calculation, and many scientific calculators have a factorial key (although it may only function correctly for integers less than 70). Alternatively the expression for P₀ can be calculated without factorials by repeated multiplication and division after cancelling common factors:

$$P_0 = \frac{22 \times 21 \times 20 \times 19 \times 18}{42 \times 41 \times 40 \times 39 \times 38} = 0.03096.$$

The probabilities for a = 1, 2, . . . , 5 can be obtained in succession. Thus,

$$P_1 = \frac{5 \times 20}{1 \times 18} \times P_0$$

$$P_2 = \frac{4 \times 19}{2 \times 19} \times P_1, \text{ etc.}$$

The results are shown above.

[Notes from JH: 1. The 5 tables from the tea-tasting experiment with to the 2x2 tables with all marginal totals = 4 are another example of this *hypergeometric* distribution]

* 2. Don't worry about the formula and the factorials; Excel has this function built in. It is called the Hypergeometric probability function, It is like the Binomial, except that instead of specifying p, one specifies the size of the POPULATION and the NUMBER OF POSITIVES IN THE POPULATION.. example, to get P1 above, one would ask for HYPGEOMDIST(a;r1;c1;N) The spreadsheet "Fisher's Exact test" uses this function; to use the spreadsheet, simply type in the 4 cell frequencies, a, b, c, and d. The spreadsheet will calculate the probability for each possible table. Then you can find the tail areas yourself.

The "Exact" Test for 2 x 2 tables continued...

This is the complete *conditional distribution* for the observed marginal totals, and the probabilities sum to unity as would be expected. Note the importance of carrying enough significant digits in the first probability to be calculated; the above calculations were carried out with more decimal places than recorded by retaining each probability in the calculator for the next stage. The observed table has a probability of 0.1253. To assess its significance we could measure the extent to which it falls into the tail of the distribution by calculating the probability of that table or of one more extreme. For a one-sided test the procedure clearly gives $P = 0.1253 + 0.0182 = 0.1435$. The result is not significant at even the 10% level.

For a two-sided test the other tail of the distribution must be taken into account, and here some ambiguity arises. Many authors advocate that the one-tailed P value should be doubled. In the present example, the one-tailed test gave $P = 0.1435$ and the two-tailed test would give $P = 0.2870$. An alternative approach is to calculate P as the total probability of tables, in either tail, which are at least as extreme as that observed in the sense of having a probability at least as small. In the present example we should have

$$P = 0.1253 + 0.0182 + 0.0310 = 0.1745.$$

The first procedure is probably to be preferred on the grounds that a significant result is interpreted as strong evidence for a difference in the *observed direction*, and there is some merit in controlling the chance probability of such a result to no more than half the two-sided significance level. The tables of Finney *et al.* (1963) enable one-sided tests at various significance levels to be made without computation provided the frequencies are not too great.

To calculate the **mid-P** value only half the probability of the observed table is included and we have

$$\text{mid-P} = 0.5(0.1253) + 0.0182 = 0.0808$$

as the one-sided value, and the two-sided value may be obtained by doubling this to give 0.1617.

The results of applying the exact test in this example may be compared with those obtained by the χ^2 test with Yates's correction. We find $X^2 = 2.39$ ($P = 0.12$) without correction and $X^2_C = 1.14$ ($P = 0.29$) with correction. The probability level of 0.29 for X^2_C agrees well with the two-sided value 0.29 from the exact test,

and the probability level of 0.12 for X^2 is a fair approximation to the exact mid-P value of 0.16.

Cochran (1954) recommends the use of the exact test, in preference to the χ^2 test with continuity correction, (i) if $N < 20$, or (ii) $20 < N < 40$ and the smallest expected value is less than 5. With modern scientific calculators and statistical software the exact test is much easier to calculate than previously and should be used for any table with an expected value less than 5.

The exact test and therefore the χ^2 test with Yates's correction for continuity have been criticized over the last 50 years on the grounds that they are conservative in the sense that a result significant at, say, the 5% level will be found in less than 5% of hypothetical repeated random samples from a population in which the null hypothesis is true. This feature was discussed in §4.7 and it was remarked that the problem was a consequence of the discrete nature of the data and causes no difficulty if the precise level of P is stated. Another source of criticism has been that the tests are conditional on the observed margins, which frequently would not all be fixed. For example, in Example 4.14 one could imagine repetitions of sampling in which 20 breast-fed infants were compared with 22 bottle-fed infants but in many of these samples the number of infants with normal teeth would differ from 5. The conditional argument is that, whatever inference can be made about the association between breast-feeding and tooth decay, it has to be made within the context that exactly five children had normal teeth. If this number had been different then the inference would have been made in this different context, but that is irrelevant to inferences that can be made when there are five children with normal teeth. Therefore, we do not accept the various arguments that have been put forward for rejecting the exact test based on consideration of possible samples with different totals in one of the margins. The issues were discussed by Yates (1984) and in the ensuing discussion, and by Barnard (1989) and Upton (1992), and we will not pursue this point further. Nevertheless, the exact test and the corrected χ^2 test have the undesirable feature that the average value of the significance level, when the null hypothesis is true, exceeds 0.5. The mid-P value avoids this problem, and so is more appropriate when combining results from several studies (see §4.7).

The "Exact" Test for 2 x 2 tables continued...

As for a single proportion, the mid-P value corresponds to an uncorrected χ^2 test, whilst the exact P value corresponds to the corrected χ^2 test. The confidence limits for the difference, ratio or odds ratio of two proportions based on the standard errors given by (4.14), (4.17) or (4.19) respectively are all approximate and the approximate values will be suspect if one or more of the frequencies in the 2 x 2 table are small. Various methods have been put forward to give improved limits but all of these involve iterations and are tedious to carry out on a calculator. The odds ratio is the easiest case. Apart from exact limits, which involve an excessive amount of calculation, the most satisfactory limits are those of Cornfield (1956); see Example 16.1 and Breslow and Day (1980, §4.3) or Fleiss (1981, §5.6). For the ratio of two proportions a method was given by Koopman (1984) and Miettinen and Nurminen (1985) which can be programmed fairly readily. The confidence interval produced gives a good approximation to the required confidence coefficient, but the two tail probabilities are unequal due to skewness. Gart and Nam (1988) gave a correction for skewness but this is tedious to calculate. For the difference of two proportions a method was given by Mee (1984) and Miettinen and Nurminen (1985). This involves more calculation than for the ratio limits, and again there could be a problem due to skewness (Gart and Nam, 1990).

Notes by JH

- The word "exact" means that the p-values are calculated using a finite discrete reference distribution -- the hypergeometric distribution (cousin of the binomial) rather than using large-sample approximations. It doesn't mean that it is the correct test. [see comment by A&B in their section dealing with Mid-P values].

While greater accuracy is always desirable, this particular test uses a 'conditional' approach that not all statisticians agree with. Moreover, compared with some unconditional competitors, the test is somewhat conservative, and thus less powerful, particularly if sample sizes are very small.

- Fisher's exact test is usually used just as a test*; if one is interested in the difference $\mu_1 - \mu_2$, the conditional approach does not yield a corresponding confidence interval for $\mu_1 - \mu_2$. [it does provide one for the comparative odds ratio parameter $\theta = \frac{1-p_1}{1-p_2} \div \frac{1-p_2}{1-p_1}$]
- Thus, one can find anomalous situations where the (conditional) test provides $P > 0.05$ making the difference 'not statistically significant', whereas the large-sample (unconditional) CI for $\mu_1 - \mu_2$, computed as $p_1 - p_2 \pm zSE(p_1 - p_2)$, does not overlap 0, and so would indicate that the difference is 'statistically significant'. [* see the Breslow and Day text Vol I, §4.2, for CI's for θ derived from the conditional distribution]
- See letter from Begin & Hanley re 1/20 mortality with pentamidine vs 5/20 with Trimethoprim-Sulfamethoxazole in patients with Pneumocystis carinii Pneumonia-Annals Int Med 106 474 1987.
- Miettinen's test-based method of forming CI's, while it can have some drawbacks, keeps the correspondence between test and CI and avoids such anomalies.
- This illustrates one important point about parameters related to binary data -- with means of interval data, we typically deal just with differences*; however, with binary data, we often switch between differences and ratios, either because the design of the study forces us to use odds ratios (case-control studies), or because the most readily available regression software uses a ratio (i.e. logistic regression for odds ratios) or because one is easier to explain than the other, or because one has a more natural interpretation (e.g. in assessing the cost per life saved of a more expensive and more efficacious management modality, it is the difference in, rather than the ratio of, mortality rates that comes into the calculation). [* the sampling variability of the estimated ratios of means of interval data is also more difficult to calculate accurately].
- Two versions of an unconditional test for the $H_0: \mu_1 = \mu_2$ are available: Liddell; Suissa and Shuster;

Abstract

Background Some claim that food sensitivities can best be identified by intradermal injection of extracts of the suspected allergens to reproduce the associated symptoms. A different dose of an offending allergen is thought to "neutralize" the reaction.

Methods To assess the validity of symptom provocation, we performed a double-blind study that was carried out in the offices of seven physicians who were proponents of this technique and experienced in its use. Eighteen patients were tested in 20 sessions (two patients were tested twice) by the same technician, using the same extracts (at the same dilutions with the same saline diluent) as those previously thought to provoke symptoms during unblinded testing. At each session three injections of extract and nine of diluent were given in random sequence. The symptoms evaluated included nasal stuffiness, dry mouth, nausea, fatigue, headache, and feelings of disorientation or depression. No patient had a history of asthma or anaphylaxis.

Results The responses of the patients to the active and control injections were indistinguishable, as was the incidence of positive responses: 27 percent of the active injections (16 of 60) were judged by the patients to be the active substance, as were 24 percent of the control injections (44 of 180). Neutralizing doses given by some of the physicians to treat the symptoms after a response were equally efficacious whether the injection was of the suspected allergen or saline. The rate of judging injections as active remained relatively constant within the experimental sessions, with no major change in the response rate due to neutralization or habituation.

Conclusions When the provocation of symptoms to identify food sensitivities is evaluated under double-blind conditions, this type of testing, as well as the treatments based on "neutralizing" such reactions, appears to lack scientific validity. The frequency of positive responses to the injected extracts appears to be the result of suggestion and chance

† Calculated according to Fisher's exact test, which assumes that the hypothesized direction of effect is the same as the direction of effect in the data. Therefore, when the effect is opposite to the hypothesis, as it is for the data below those of Patient 9, the P value computed is testing the null hypothesis that the results obtained were due to change as compared with the possibility that the patients were more likely to judge a placebo injection as active than an active injection.

Table 1: Responses of 18 Patients Forced to Decide Whether Injections Contained an Active Ingredient or Placebo

Pt. No*	Active Injection		Placebo Injection		P Value†
	resp	no resp	resp	no resp	
3	2	1	1	8	0.13
1	2	1	2	7	0.24
14a	2	1	2	7	0.24
12	1	2	0	9	0.25
16	2	1	3	6	0.36
18	2	1	4	5	0.50
14b	1	2	2	7	0.87
4	1	2	2	7	0.87
5	1	2	2	7	0.87
9	0	3	0	9	--
2a	0	3	1	8	0.75
13	0	3	1	8	0.75
15	1	2	3	6	0.76
6	0	3	2	7	0.55
8	0	3	2	7	0.55
17	1	2	5	4	0.50
2b	0	3	3	6	0.38
7	0	3	3	6	0.38
10	0	3	3	6	0.38
11	0	3	3	6	0.38

*Patients were numbered in the order they were studied.

The order in the table is related to the degree that the results agree with the hypothesis that patients could distinguish active injections from placebo injections. The results listed below those of Patient 9 do not support this hypothesis, placebo injections were identified as active at a higher rate than were true active injections. The letters a and b denote the first and second testing sessions, respectively, in Patients 2 and 14. true active injections. ID denotes intradermal, and SC subcutaneous.

The value is the P value associated with the test of whether the common odds ratio (the odds ratio for all patients) is equal to 1.0. The common odds ratio was equal to 1.13 (computed according to the Mantel-Haenszel test).

Notes on P-Values from Fisher's Exact Test in previous article

		Response		Total
		+	-	
<u>Patient no. 3</u>	Active Injection	2	1	3
	Placebo Injection	1	8	9

		3	9	

All possible tables with a total of 3 +ve responses

	0 3	1 2	2 1	3 0
	3 6	2 7	1 8	0 9
prob	9 • 8 • 7	3 • 3	2 • 2	1 • 1
	-----	• ---	• ---	• ---
	12 • 11 • 10	1 • 7	2 • 8	3 • 9
	<u>0.382</u>	<u>0.491</u>	<u>0.123</u>	<u>0.005</u>
(pt #)	(2b, 7, 10, 11)	(14b, 4, 5)	(3)	
P-Value*	<u>1.0</u>	<u>0.618</u>	<u>0.128</u>	<u>0.005</u>

		Response		Total
		+	-	
<u>Patient no. 1</u>	Active Injection	2	1	3
	Placebo Injection	2	7	9

		4	8	

All possible tables with a total of 4 +ve responses

	0 3	1 2	2 1	3 0
	4 5	3 6	2 7	1 8
prob	8 • 7 • 6	3 • 4	2 • 3	1 • 2
	-----	• ---	• ---	• ---
	12 • 11 • 10	1 • 6	2 • 7	3 • 8
	<u>0.255</u>	<u>0.510</u>	<u>0.218</u>	<u>0.018</u>
(pt #)	(15)	(1, 14a)		
P-Value	<u>1.0</u>	<u>0.745</u>	<u>0.236</u>	<u>0.018</u>
(*1-sided, guided by H _{alt} : of +ve responses with Active > of +ve responses with Placebo)				

		Response		Total
		+	-	
<u>Patient no. 18</u>	Active Injection	2	1	3
	Placebo Injection	4	5	9

		6	6	

All possible tables with a total of 6 +ve responses

	0 3	1 2	2 1	3 0
	6 3	5 4	4 5	3 6
prob	6 • 5 • 4	3 • 6	2 • 5	1 • 4
	-----	• ---	• ---	• ---
	12 • 11 • 10	1 • 4	2 • 5	3 • 6
	<u>0.091</u>	<u>0.409</u>	<u>0.409</u>	<u>0.091</u>
(pt #)		(17)	(18)	
P-Value	<u>1.0</u>	<u>0.909</u>	<u>0.500</u>	<u>0.091</u>
(1-sided, as above)				

In Table 1, the P-values for patients below patient 9 are calculated as 1-sided, but guided by the opposite H_{alt} from that used for the patients in the upper half of the table, i.e. by H_{alt}:

π of +ve responses with Active < π of +ve responses with Placebo.

It appears that the authors decided the "sidedness" of the H_{alt} after observing the data!!! and that they used different H_{alt} for different patients!!!

The Namibian government expelled the authors from Namibia following the publication of this article; the reason given was that their "data and conclusions were premature" .. jh ¶

Since 1900 the world's population has increased from about 1.6 (to over 5 billion) the U.S. population has kept pace, growing from nearly 75 to 260 million. While the expansion of humans and environmental alterations go hand in hand, it remains uncertain whether conservation programs will slow our biotic losses. Current strategies focus on solutions to problems associated with diminishing and less continuous habitats, but in the past, when habitat loss was not the issue, active intervention prevented extirpation. Here we briefly summarize intervention measures and focus on tactics for species with economically valuable body parts, particularly on the merits and pitfalls of biological strategies tried for Africa's most endangered pachyderms, rhinoceroses.

[...]

Given the inadequacies of protective legislation and enforcement, Namibia, Zimbabwe, and Swaziland are using a controversial preemptive measure, dehorning (Fig. D) with the hope that complete devaluation will buy time for implementing other protective measures (7) In Namibia and Zimbabwe, two species, black and white rhinos (*Ceratotherium simum*), are dehorned, a tactic resulting in *sociological and biological uncertainty: Is poaching deterred? Can hornless mothers defend calves from dangerous predators?*

On the basis of our work in Namibia during the last 3 years (8) and comparative information from Zimbabwe, some data are available. Horns regenerate rapidly, about 8.7 cm per animal per year, so that 1 year after dehorning the regrown mass exceeds 0.5 kg. Because poachers apparently do not prefer animals with more massive horns (8), frequent and costly horn removal may be required (9). In Zimbabwe, a population of 100 white rhinos, with at least 80 dehorned, was reduced to less than 5 animals in 18 months (10). These discouraging results suggest that intervention by itself is unlikely to eliminate the incentive for poaching. Nevertheless, some benefits accrue when governments, rather than poachers, practice horn harvesting, since less horn enters the black market Whether horn stockpiles may be used to enhance conservation remains controversial, but mortality risks associated with anesthesia during dehorning are low (5).

Biologically, there have also been problems. Despite media attention and a bevy of allegations about the soundness of dehorning (11), serious attempts to determine whether dehorning is harmful have been remiss. A lack of negative effects has been suggested because (i) horned and dehorned individuals have interacted without subsequent injury; (ii) dehorned animals have thwarted the advance of dangerous predators; (iii) feeding is normal; and (iv) dehorned mothers have given birth (12) However, most claims are anecdotal and mean little without attendant data on demographic effects. For instance, while some dehorned females give birth, it may be that these females were pregnant when first immobilized. Perhaps others have not conceived or have lost calves after birth. Without knowing more about the frequency of mortality, it seems premature to argue that dehorning is effective.

We gathered data on more than 40 known horned and hornless black rhinos in the presence and absence of dangerous carnivores in a 7,000 km² area of the northern Namib Desert and on 60 horned animals in the 22,000 km² Etosha National Park. On the basis of over 200 witnessed interactions between horned rhinos and spotted hyenas (*Crocuta crocuta*) and lions (*Panthera leo*) we saw no cases of predation, although mothers charged predators in about 45% of the cases. Serious interspecific aggression is not uncommon elsewhere in Africa, and calves missing ears and tails have been observed from South Africa, Kenya, Tanzania, and Namibia (13).

To evaluate the vulnerability of dehorned rhinos to potential predators, we developed an experimental design using three regions:

• Area A had horned animals with spotted hyenas and occasional lions

• Area B had dehorned animals lacking dangerous predators,

• Area C consisted of dehorned animals that were sympatric with hyenas only.

Populations were discrete and inhabited similar xeric landscapes that averaged less than 125 mm of precipitation annually. Area A occurred north of a country-long veterinary cordon fence, whereas animals from areas B and C occurred to the south or east, and no individuals moved between regions.

The differences in calf survivorship were remarkable. All three calves in area C died within 1 year of birth, whereas all calves survived for both dehorned females living without dangerous predators (area B; n = 3) and for horned mothers in area A (n = 4). Despite admittedly restricted samples, the differences are striking [Fisher's (3 x 2) exact test, P = 0.017; area B versus C, P = 0.05; area A versus C, P = 0.0291 ††. The data offer a first assessment of an empirically derived relation between horns and recruitment.

Our results imply that hyena predation was responsible for calf deaths, but other explanations are possible. If drought affected one area to a larger extent than the others, then calves might be more susceptible to early mortality. This possibility appears unlikely because all of western Namibia has been experiencing drought and, on average, the desert rhinos in one area were in no poorer bodily condition than those in another. Also, the mothers who

lost calves were between 15 to 25 years old, suggesting that they were not first time, inexperienced mothers (14). What seems more likely is that the drought-induced migration of more than 85% of the large, herbivore biomass (kudu, springbok, zebra, gemsbok, giraffe, and ostrich) resulted in hyenas preying on an alternative food, rhino neonates, when mothers with regenerating horns could not protect them.

Clearly, unpredictable events, including drought, may not be anticipated on a short-term basis. Similarly, it may not be possible to predict when governments can no longer fund antipoaching measures, an event that may have led to the collapse of Zimbabwe's dehorned white rhinos. Nevertheless, any effective conservation actions must account for uncertainty. In the case of dehorning, additional precautions must be taken. [...]

	A	B	C
survived	4	3	0
died	0	0	3
	4	3	3

B vs C

	B	C	B	C	B	C	tot*
survived	3	0	2	1	1	2	0 3 3
died	0	3	1	2	2	1	3 3 3
	3	3	3	3	3	3	



A vs C

	A	C	A	C	A	C	tot*
survived	4	0	3	1	2	2	1 3 4
died	0	3	1	2	2	1	3 3 3
	4	3	4	3	4	3	
prob	1/35	12/35	18/35	4/35			

¶ Do you agree?