Contents

1 Comparative measures / parameters:

Measure	Comparative Parameter	Estimate	New Scale
(Risk or Prevalence) **Difference**	$\pi_1 - \pi_2$	$p_1 - p_2$	
(Risk or **NNT**	$1/\{\pi_1 - \pi_2\}$	$1/\{p_1 - p_2\}$	Number Needed to Treat
(Risk or Prevalence) **Ratio**	$\frac{\pi_1}{\pi_2}$	$\frac{p_1}{p_2}$	$\log \frac{p_1}{p_2} = \log p_1 - \log p_2$
Odds Ratio	$\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$	$\frac{odds_1}{odds_2}$	$log[\frac{odds_1}{odds_2}] = logit_1 - logit_2$

Cf. Rothman 2002 p. 135 Eqns 7-2, 7-3 and 7-6.

2 Large-sample CI for Comparative Parameter (if 2 component estimates are uncorrelated)

2.1 In General: (if work in new scale, must back-transform)

$$\begin{array}{lcl} estimate_1 - estimate_2 & \pm & z \times \mathrm{SE}[estimate_1 - estimate_2] \\ estimate_1 - estimate_2 & \pm & z \times (\mathrm{Var}[estimate_1] + \mathrm{Var}[estimate_2])^{1/2}. \end{array}$$

2.2 In Particular

Risk/Prevalence Difference

$$\begin{aligned} p_1 - p_2 \pm z \times SE[p_1 - p_2] &= p_1 - p_2 \pm z \times (SE^2[p_1] + SE^2[p_2])^{/2} \\ &= p_1 - p_2 \pm z \times (p_1 q_1/n_1 + p_2 q_2/n_2)^{1/2} \end{aligned}$$

Risk/Prevalence Ratio

$$\text{antilog } \{\log(p_1/p_2) \pm z \times (SE^2[\log p_1] + SE^2[\log p_2])^{1/2}\},$$

where, for $i = 1, 2$,

$$SE^2[\log p_i] = Var[\log p_i] = 1/\#positive_i - 1/\#total_i.$$

Odds ratio[1]

$$\text{antilog } \{\log[oddsratio] \pm z \times (SE^2[logit_1] + SE^2[logit_2])^{1/2}\}$$

where, for $i = 1, 2$,

$$SE^2[logit_i] = Var[logit_i] = 1/\#positive_i + 1/\#negative_i.$$

$\mathrm{Var}[\log or] = \underline{1/a + 1/b} + \underline{1/c + 1/d}$ for CI_{OR} → "Woolf's Method."

2.3 Large-sample test of $\pi_1 = \pi_2$

Equivalent to test of $\pi_1 - \pi_1 = 0 \rightarrow$ Risk or Prevalence Difference = 0.

$\pi_1/\pi_2 = 1 \rightarrow$ Risk or Prevalence Ratio = 1.

$\frac{pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = 1 \rightarrow$ Odds Ratio = 1.

$$\begin{aligned} z &= (p_1 - p_2 - \{\Delta = 0\}) \,/\, SE[p_1 - p_2] \\ &= (p_1 - p_2) \,/\, (p[1-p]/n_1 + p[1-p]/n_2)^{1/2} \end{aligned}$$

where $p = y/n$, with $y = y_1 + y_2$; $n = n_1 + n_2$.

[1] The Odds Ratio (OR) is close to the Risk Ratio when the 'denominator' odds is low, e.g. under 0.1, and the Risk Ratio is not extreme. For example, if $\pi_1 = 0.16$, and $\pi_2 = 0.08$, so that the Risk Ratio is 2, then OR = (0.16/0.84)/(0.08/0.92) = 2.2; but the approximation worsens with increasing π_2 and increasing Risk Ratio.

Examples:

0 **The generic 2×2 contingency table:**

	+	−	All
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 primary infertility:**[2]

	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:**[3]

	No cold	≥ 1 cold	Total subjects
Vitamin C	105(26%)	302	407(100%)
Placebo	76(18%)	335	411(100%)
Total	181(22%)	637	818(100%)

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

	Indep't.	Dep'nt	Total no. pts
Stroke Unit	67(66%)	34	101(100%)
Medical Unit	46(51%)	45	91 (100%)
Total	113(59%)	79	192(100%)

Worked example: Stroke Unit vs. Medical Unit

95% CI for $\Delta\pi$:

$$\begin{aligned} & 0.66 - 0.51 \pm z \times (0.66 \times 0.34/101 + 0.51 \times 0.49/91)^{/2} \\ = \; & 0.15 \pm 1.96 \times 0.07 \\ = \; & 0.15 \pm 0.14. \end{aligned}$$

Test $\Delta\pi = 0$: [carrying several decimal places, for comparison with χ^2]

$$\begin{aligned} z \; = \; & (0.6634 - 0.5054) \;/|;(0.5885 \times 0.4115 \times \{1/101 + 1/91\})^{1/2} \\ = \; & 0.1580 \,/\, 0.0711 \\ = \; & 2.22 \;\rightarrow\; P = 0.026 \text{ (2-sided).} \end{aligned}$$

Worked example: Vitamin C and the common cold

95% CI for $\Delta\pi$:

$$\begin{aligned} & 0.26 - 0.18 \pm z \times (0.26 \times 0.74/407 + 0.18 \times 0.81/411)^{/2} \\ = \; & 0.18 \pm 1.96 \times 0.03 \\ = \; & 0.18 \pm 0.06. \end{aligned}$$

Test $\Delta\pi = 0$:

$$\begin{aligned} z \; = \; & (0.258 - 0.185) \;/|;(0.221 \times 0.779 \times \{1/407 + 1/411\})^{1/2} \\ = \; & 0.073 \,/\, 0.029 \\ = \; & 2.52 \;\rightarrow\; P = 0.006 \text{ (1-sided) or } 0.012 \text{ (2-sided).} \end{aligned}$$

2.4 CI for Risk Ratio (a.k.a. Relative Risk) or Prevalence Ratio cf. Rothman2002 p.135

Example: Vitamin C and the common cold ... Revisited

	No cold	≥ 1 cold	Total no. subjects
Vitamin C	105(26%)	302(74%)	407(100%)
Placebo	76(18%)	335(82%)	411(100%)
Total	181(22%)	637	818(100%)

$$\widehat{RR} = \frac{Prob[\geq 1 \text{ cold} \mid \text{Vitamin C}\,]}{Prob[\geq 1 \text{ cold} \mid \quad \text{Placebo}\,]} = \frac{74\%}{82\%} = 0.91$$

CI[RR]:

$$\begin{aligned} & antilog\{\log 0.91 \pm z \times SE[\log p_1 - \log p_2]]\} \\ = \;& antilog\{\log 0.91 \pm z \times (SE^2[\log p_1] + SE^2[\log p_2])^{1/2}\}. \end{aligned}$$

$$\begin{aligned} SE^2[\log p_1] = Var[\log p_1] = 1/302 - 1/407 = 0.000854; \\ SE^2[\log p_2] = Var[\log p_2] = 1/335 - 1/411 = 0.000552. \end{aligned}$$

So, CI[RR]:

$$\begin{aligned} & antilog\{\log 0.91 \pm z \times (0.000854 + 0.000552)^{1/2}\} \\ = \;& antilog\{\log 0.91 \pm 0.073\} = 0.85 \text{ to } 0.98. \end{aligned}$$

Shortcut:

Calculate $\exp\{z \times SE[\log \widehat{RR}]\}$ and use it as a multiplier and divider of $\widehat{RR}$.

In our e.g., $\exp\{z \times SE[\log \widehat{RR}]\} = \exp\{0.073\} = 1.076$.

Thus $\{RR_{LOWER}, RR_{UPPER}\} = \{0.91 \div 1.076, 0.91 \times 1.076\} = \{0.85 \text{ to } 0.98\}$.

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.

```
SAS                               Stata

PROC FORMAT;                      Immediate:  csi 302 335 105 76
VALUE onefirst 0="z0"             cs stands for 'cohort study'
1="a1";
DATA CI_RR_OR;                    input vitc cold npeople
INPUT vitC cold npeople;
LINES;
1 1 302                           1 1 302
1 0 105                           1 0 105
0 1 335                           0 1 335
0 0 76                            0 0 76
;                                 end
PROC FREQ data=CI_RR_OR
ORDER=FORMATTED;                  cs cold vitc [freq=npeople]
TABLES vitC*cold / CMH;
WEIGHT npeople;
FORMAT vitC cold onefirst;
RUN;
```

2.5 CI for Odds Ratio **cf. Rothman 2002 p. 139**

	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**
odds **Ratio** = $\frac{2.88}{4.41}$	= 0.65 → $\widehat{OR} = 0.65$	

$CI[OR] = antilog\ \{\log[oddsRatio] \pm z\ SE[logit_1 - logit_2]\ \}$

$$SE^2[logit_1] = \frac{1}{\#positive_1} + \frac{1}{\#negative_1}$$

$$SE^2[logit_1] = \frac{1}{\#positive_2} + \frac{1}{\#negative_2}$$

$$SE[logit_1 - logit_2] = \left\{ \left(\frac{1}{302} + \frac{1}{105} \right) + \left(\frac{1}{335} + \frac{1}{76} \right) \right\}^{1/2} = 0.17$$

$$z \times SE[logit_1 - logit_2] = 1.96 \times 0.17 = 0.33$$

$$antilog\ \{\log 0.65 \pm 0.33\ \} = \exp\ \{-0.43 \pm 0.33\} = 0.47 to 0.90$$

`From SAS`

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 it data come from a 'case-control' or 'cohort' study.

```
From Stata
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
```

3 “Test-based CI’s”

3.1 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 starum) 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.

3.2 “Test-based” CI’s ... specific applications

- **Difference of 2 proportions** $\pi_1 - \pi_2$ (Risk or Prevalence Difference)

  Observe: p_1 and p_2 and (maybe via p-value) the calculated value of X^2 This implies that

  $$(observed\ X^2\ value)^{1/2} = observed\ X\ value = observed\ z\ value;$$

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

  So... $SE[p_1 - p_2] = (p_1 - p_2)\ /\ observed\ z\ statistic \qquad \{use\ +\text{ve}\ sign\}$

  95% CI for $p_1 - p_2$:

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

  i.e. ...

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

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

  $$(p_1 - p_2)\left\{1 \mp \frac{z\ value\ for\ 95\%\ CI}{observed\ z\ statistic}\right\} \qquad (1a)$$

or, in terms of a reported chi-squared statistic

$$(p_1 - p_2)\left\{1 \mp \frac{z\ value\ for\ 95\%\,CI}{Sqrt[observed\ chi - squared\ statistic]}\right\}. \qquad (1b)$$

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.

- **Ratio** of 2 proportions $\pi_1\ /\ \pi_2$
  (Risk Ratio; Prevalence Ratio; Relative Risk; "RR")

  Observe:

  1. $rr = p_1\ /\ p_2$ and
  2. (maybe via p-value) the value of X^2 statistic (H_0: RR = 1)
     $\rightarrow (observed\, X^2 value)^{1/2} = observed\, X\ value = observed\, z\ value.$

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

  $$observed\ z\ value = \frac{\log rr - 0}{SE[\log rr]}$$

  This implies that

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

  95% CI for $\log RR$:

  $$\log rr \mp \{z\ value\ for\ 95\%\ CI\} \times SE[\log rr]$$

  i.e. ...

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

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

  $$log[rr] \times \left\{1 \pm \frac{z\ value\ for\ 95\%\ CI}{observed\ z\ statistic}\right\} \qquad (2a)$$

  Going back to RR scale, by taking antilogs[4]...

  95% CI for RR:

  $$rr\ to\ power\ of\ \left\{1 \pm \frac{z\ value\ for\ 95\%}{observed\ z\ statistic}\right\} \qquad (2b)$$

- **Ratio** of 2 odds $\pi_1/(1-\pi_1)$ and $\pi_2/(1-\pi_2)$ (Odds Ratio; "OR")

  Observe:

  1. $or = \frac{p_1/(1-p_2)}{p_2/(1-p_2)} = \frac{ad}{bc}$ and
  2. (maybe via p-value) the value of X^2 statistic (H_0: OR = 1)
     $\rightarrow (observed\ X^2 value)^{1/2} = observed\ X\ value = observed\ z\ value$

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

  $$observed\ z\ value\ \ = \frac{\log or - 0}{SE[\log or]}$$

  This implies that

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

  95% CI for $\log OR$:

  $$\log or \mp \{z\ value\ for\ 95\%\ CI\} \times SE[\log or] \qquad (3a)$$

  i.e. ...

  $$\log or \pm \{z\ value\ for\ 95\%\ CI\} \times \frac{\log or}{observed\ z\ value} \qquad (3b)$$

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

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

  Going back to OR scale, by taking antilogs[5]...

  95% CI for OR:

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

See Section 13.3 of Miettinen's "Theoretical Epidemiology"

[4] $antilog[log[a]\infty b] = \exp[log[a]\infty b] = \{\exp[log[a]]\}$ to power of $b = a$ to power of b

4 Sample Size considerations...

4.0.1 CI for $\pi_1 - \pi_2$

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

- large-sample CI: $p_1 - p_2 \pm Z\, SE[p_1 - p_2] = p_1 - p_2 \pm ME$
- $SE[p_1 - p_2] = \{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2\}^{1/2}$.

  Simplify (involves some approximation) by using an average p.

  If use equal n's, then

  $$n\ per\ group = \frac{2 \times p(1-p) \times Z^2_{\alpha/2}}{ME^2}$$

  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

  $$[max]\ n\ per\ group = \frac{\frac{1}{2}\, Z^2_{\alpha/2}}{ME^2}$$

4.0.2 Test involving π_T and π_C

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

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

n per group

$$\begin{aligned} &= \frac{\{Z_{\alpha/2}\sqrt{2\pi_C\{1-\pi_C\}} - Z_\beta\sqrt{\pi_C\{1-\pi_C\} + \pi_T\{1-\pi_T\}}\}^2}{\Delta^2} \\ &\approx 2(Z_{\alpha/2} - Z_\beta)^2\left\{\frac{\bar{\pi}(1-\bar{\pi})}{\Delta^2}\right\} \\ &= 2\{Z_{\alpha/2} - Z_\beta\}^2\left\{\frac{\sigma_{0,1}}{\Delta}\right\}^2 \qquad (4) \end{aligned}$$

If $\alpha = 0.05(2-sided)$ & $\beta = 0.2 ... Z_\alpha = 1.96$; $Z_\beta = -0.84$, then $2(Z_{\alpha/2} - Z_\beta)^2 = 2\{1.96 - (-0.84)\}^2 \approx 16$, i.e. $n\,per\,group \approx 16 \times \frac{\bar{\pi}\{1-\bar{\pi}\}}{\Delta^2}$.

$\rightarrow$ $n_T \approx 100$ & $n_C \approx 100$ if $\pi_T = 0.6$ & $\pi_C = 0.4$.

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: See Notes on Sample Size Calculations for Inferences Concerning Means.

4.0.3 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] = α.

Work in $\log or$ scale; $SE[\log or] = (1/a + 1/b + 1/c + 1/d)^{1/2}$.

Need

$$Z_{\alpha/2}\, SE_0[\log or] + Z_\beta SE_{alt}[\log or] < \Delta.$$

where

$$\Delta = \log[OR_{alt}]$$

Substitute expected $a, |; b,\ c,\ d$ values under null and alt. into SE's and solve for number of cases and controls.

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

Key Points: $\log or$ 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

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

   for various values of "multiple" [cf earlier notes "effect of unequal sample sizes"]

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

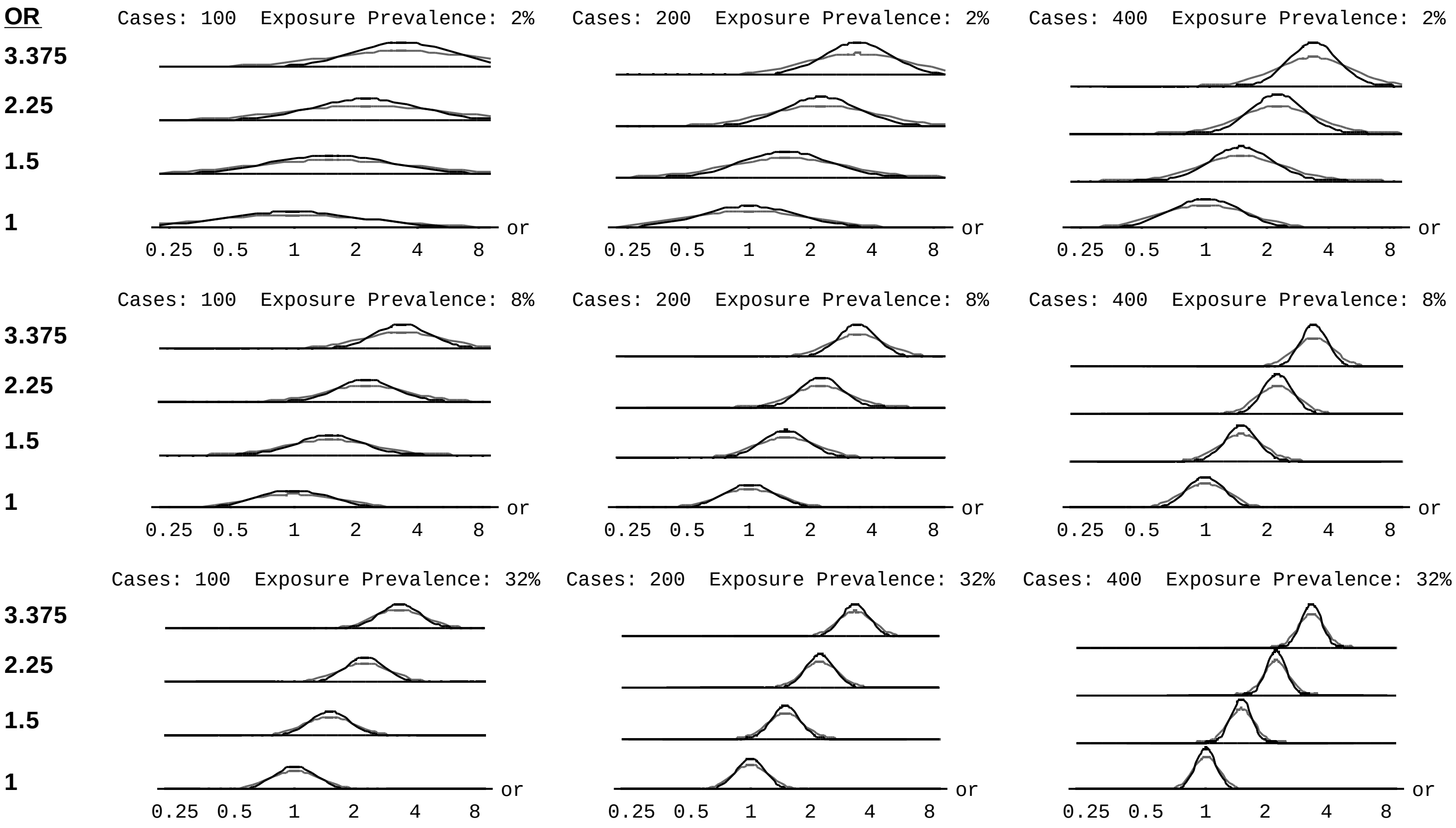
   Examine the functions

   $$1/(\text{no. of exposed cases}) + 1/(\text{no. of unexposed cases})$$

   and

   $$1/(\text{no. of exposed controls}) + 1/(\text{no. of unexposed controls}).$$

Factors affecting variability of estimates from, and statistical[5] power of, case-control studies



jh 1995-2003

Reading graphs: (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 $\alpha = 0.05$ (2-sided) 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.

[5] **Power larger if ...** 1. non-null OR >> 1 (cf. 2.5 vs 2.25 vs 3.375); 2 exposure common (cf. 2% vs 8% vs 32%) and not near universal; 3 use more cases (cf. 100 vs. 200 vs. 400), and controls/case (1 vs 4).

5 Small sample methods:

Test:

Since a risk difference of zero implies a risk ratio, or odds ratio, of 1, all three can be tested in the same way.

U (unconditional)
Suissa S; Shuster JJ. Exact Unconditional Sample Sizes for the 2×2 Binomial Trial; Journal of the Royal Statistical Society. Series A (General) Vol. 148, No. 4 (1985), pp. 317-327.

C (conditional)
Fisher 1935, JRSS Vol 98, p 48. (central) Hypergeometric distribution, obtained by conditioning on (treating as fixed) all marginal frequencies.

Confidence Interval:

5.1 Risk Difference

See section 3.1.2 of Sahai and Khurshid (1996).

5.2 Risk Ratio

See section 3.1.2 of Sahai and Khurshid (1996).

5.3 Odds Ratio: Point- and Interval-estimation

See section 4.1.2 of Sahai and Khurshid (1996), and Chapter of Volume I of Breslow and Day. See also example 1, pp 48-51, in Fisher 1935.

Elaboration on equation 4.11 in Sahai and Khurshid , and on the (what we now call the *non-central* hypergeometric random variable whose distribution is given in the middle of p 50 of Fisher's article.

Let $Y_i \sim \text{Binomial}(n_i, \pi_i)$, $i = 1, 2$, be 2 independent binomial random variables.

We wish to make inference regarding the parameter

$$\psi = \{\pi_1/(1-\pi_1\}/\{\pi_2/(1-\pi_2\}.$$

We can do so by considering only those data configurations which have the same total number of 'positives', $y_1 + y_2 = y$, say, as were observed in the actual study, and then considering the distribution of $Y_1 \mid y$.

$$Prob[Y_1 = y_1\,;\, Y_2 = y_2] = \ {}^{n_1}\mathrm{C}_{y_1}\, \pi_1^{y_1}(1-\pi_1)^{n_1-y_1} \times {}^{n_2}\mathrm{C}_{y_2}\, \pi_2^{y_2}(1-\pi_2)^{n_2-y_2}.$$

If we condition on $Y_1 + Y_2 = y$, then

$$Prob[Y_1 = y_1 \mid Y_1 + Y_2 = y] = Prob[Y_1 = y_1\,;\, Y_2 = y - y_1]/Prob[Y_1 + Y_1 = y].$$

If we rewrite the quantity

$$\pi_1^{y_1}(1-\pi_1)^{n_1-y_1} \times \pi_2^{y_2}(1-\pi_2)^{n_2-y_2}$$

as

$$\pi_1^{y_1}(1-\pi_1)^{-y_1}\pi_2^{-y_2}(1-\pi_2)^{y_1} \times (1-\pi_1)^{n_1}\pi_2^{y}(1-\pi_2)^{n-y}$$

we see that it simplifies to

$$\psi^{y_1} \times (1-\pi_1)^{n_1}\, \pi_2^{y}\, (1-\pi_2)^{n-y}$$

and that the last three terms do not involve ψ and do not involve the random variable y_1. Since they appear in both the numerator and the denominator of the conditional probability, they cancel out.

This we can write the conditional probability $Prob[Y_1 = y_1 \mid Y_1 + Y_2 = y]$ as

$$Prob[\,y_1 \mid y\,] = \ {}^{n_1}\mathrm{C}_{y_1}\ {}^{n_2}\mathrm{C}_{y-y_1}\, \psi^{\,y_1} \,/\, \Sigma\ {}^{n_1}\mathrm{C}_{y_1'}\ {}^{n_2}\mathrm{C}_{n-y_1'}\, \psi^{\,y_1'},$$

where the summation is over those y_1' values that are compatible with the 4 marginal frequencies.

Aside: you will note that if we set $\psi = 1$, the probabilities are the same as those in the central hypergeometric distribution, used for Fisher's exact test of two binomial proportions. Indeed, Fisher, in page 48-49 of his 1935 paper, first computes the null probabilities for the 2×2 table.

Conviction of Like-sex Twins of Criminals

	Convicted.	Not Convicted.	Total.
Monozygotic	10(a)	3(b)	13
Dizygotic	2(c)	15(d)	17
Total	12	18	30

[We use y_1 and y_2 where epidemiologists typically use a and c.]

He calculated that the probability that 1, 2, 3,... monozygotic twins would escape conviction[6] was (1/6 652 325)×{1, 102, 2992, ...}. Thus, "a discrepancy from proportionality as great or greater than that observed, will arise, subject to the conditions specified by the ancillary information, in exactly 3,095 trials out of 6,652,325 or approximately once in 2,150 trials."

He then went on to work out the lower limit of the 90% 2-sided CI (or a 95% 1-sided CI), for the odds ratio: i.e. for the odds, $\pi_{mono-z}/(1-\pi_{mono-z})$, of criminals to non-criminals in twins of monozygotic criminals divided by the corresponding odds $\pi_{di-z}/(1-\pi_{di-z})$, in twins of dizygotic criminals.

Let Y_{mono} be the number of MZ twins convicted. Fisher finds the value ψ_L such that

$$Prob[\,Y_{mono} \geq 10 \mid \psi_L\ ,\ y = 12\,] = 0.05.$$

He reports that this value is $1/0.28496 \approx 3.509$. In the Excel spreadsheet for Fisher's exact test and exact CI for OR (on website), you can verify that indeed, with $\psi_L = 3.509$, $Prob[\,Y_{mono} \geq 10 \mid \psi = 3.509\ ,\ y = 12\,] = 0.05$.

One has to admire Fisher's ability, in 1935, to solve a polynomial equation of order 12, namely

$$\frac{1 + 102\psi + 2992\psi^2}{1 + 102\psi + 2992\psi^2 + \cdots + 476\psi^{12}} = 0.05.$$

5.3.1 Point estimation of ψ under Hypergeometric Model

See section x.x of Breslow and Day, Volume I.

It will come as a surprise to many that *there are 2 point estimators of* ψ:

one, the familiar – *unconditional* – based on the "2 independent Binomials" model, with two random variables y_1 and y_2, and

the other – *conditional* – based on the *single* random variable $y_1 \mid y$ with a Non-Central Hypergeometric distribution.

While the two estimators yield similar estimates when sample sizes are large, the estimates can be quite different from each other in small sample situations.

Estimator, based on Unconditional Approach:

The estimator derives from the principle that if there are two parameters θ_1 and θ_2, with Maximum Likelihood Estimators $\hat{\theta_1}$ and $\hat{\theta_2}$, then the Maximum Likelihood Estimator of θ_1/θ_2 is $\hat{\theta_1}/\hat{\theta_2}$.

Thus, since $\hat{\pi}_1 = 10/13$, and $\hat{\pi}_2 = 2/17$, we have

$$\hat{\psi}_{UMLE} = \frac{(10/13)/(2/13)}{(2/17)/(15/17)} = \frac{10 \times 15}{3 \times 2} = 25 = \frac{a \times d}{b \times c}.$$

Estimator, based on Conditional Approach:

The Maximum Likelihood Estimate $\hat{\psi}_{CMLE}$ is the solution of $d \log L/d\psi = 0$.

If we use Σ as shorthand for the denominator of $prob[\,y_1 \mid y\,]$, then $\hat{\psi}_{CMLE}$ is the solution of

$$\frac{y_1}{\psi} = \frac{d \log \Sigma}{d\psi} = \frac{d\Sigma}{d\psi} \times \frac{1}{\Sigma}.$$

Re-arranging, we find that $\hat{\psi}_{CMLE}$ is the solution of

$$y_1 = \mathrm{E}[\,Y_1 \mid \psi\,].$$

In this case the CMLE of ψ is the same as the estimate obtained by equating the observed and expected moment (the "Method of Moments").

Using the same spreadsheet used above, we find that the value of ψ that satisfies this estimating equation is

$$\hat{\psi}_{CMLE} = 21.3.$$

It can be shown that, in any given dataset, $\hat{\psi}_{CMLE}$ is *closer to the null* (i.e., to $\psi = 1$) than the $\hat{\psi}_{MLE}$ is. Indeed, it the CMLE can be can be seen as a UMLE that has been shrunk towards the null.[7]

[8]

[6] the range is 1 to 13; 0 cannot escape, since then there would be 13 convicted in the first row, but there are only 12 convicted in all.

[7] See Hanley JA, Miettinen OS. An Unconditional-like Structure for the Conditional Estimator of Odds Ratio from 2 x 2 Tables. Biometrical Journal 48 (2006) 1, 2334 DOI: 10.1002/bimj.200510167

[8] [Notes from JH]:

- The 5 tables from the tea-tasting experiment with the 2x2 tables with all marginal totals = 4 are another example of this hypergeometric distribution
- Excel has 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 P_1 above, one would ask for HYPERGEODIST(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. You can also use it for the non-null (non-central) hypergeometric distribution.

5.4 The "Exact" Test for 2 x 2 tables

5.4.1 Material taken from Armitage & Berry §4.9. Material on hand-calculation of null probabilities is omitted

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

$$P[a|r_1, r_2, c_1, c_2] = \frac{r_1!r_2!r_3!r_4!}{N!a!b!c!d!} \quad (5)$$

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 he difficulties of interpretation which may arise. The data in Table 4.6, due to M. Hellman, are discussed by Yates (1934).

Table 1: Data on malocclusion of teeth in infants (Yates, 1934)

	Infants with		
	Normal teeth	Malocclusion	Total
Breast-fed	4	16	20
Bottle-fed	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 2 Below them are shown the probabilities of these tables, calculated under the null hypothesis.

Table 2: Cell frequencies in tables with the same marginal totals as those in Table 1

	0	20	20	1	19	20	2	18	20	3	17	20	4	16	20
	5	17	22	4	18	22	3	19	22	2	20	22	1	21	22
	5	37	42	5	37	42	5	37	42	5	37	42	5	37	42
a	0			1			2			3			4		
P_a	0.1720			0.3440			0.3096			0.1253			0.0182		

Table 2 continued ...

	5	15	20
	0	22	22
	5	37	42
a	5		
P_a	0.0182		

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 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 X^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).

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 $\Delta = \pi_1 \pi_2$, the conditional approach does not yield a corresponding confidence interval for Δ. [it does provide one for the comparative odds ratio parameter $\psi = \frac{\pi_1/(1-\pi_1}{\pi_2/(1-\pi_2}$.

- 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 Δ, computed as $p_1 - p_2 \pm z \; SE(p_1 - p_2)$, 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 ψ 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 Preumonia-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 that 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].

6 (Mis-)Application; Costly Application

6.1 Fisher's Exact Test in a Double-Blind study of Symptom Provocation to Determine Food Sensitivity (N Engl J Med 1990; 323: 429-33)

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.

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 above article

Patient number 3:

	Response		Total
	+	-	
Active Injection	2	1	3
Placebo Injection	1	8	9
	3	9	

All possible tables with a total of 3 +ve responses

	0 3 3 6	1 2 2 7	2 1 1 8	3 0 0 9
Prob	$\frac{9\times8\times7}{12\times11\times10}$ $= 0.382$	$0.382 \times \frac{3\times3}{1\times7}$ $= 0.491$	$0.491 \times \frac{2\times2}{2\times8}$ $= 0.123$	$0.123 \times \frac{1\times1}{3\times9}$ $= 0.005$
(pt #)	(2b, 7, 10, 11)	(14b, 4, 5)	(3)	
P-Value*	1.0	0.618	0.128	0.005

Patient number 1:

	Response		Total
	+	-	
Active Injection	2	1	3
Placebo Injection	2	7	9
	4	8	

All possible tables with a total of 4 +ve responses

	0 3 4 5	1 2 3 6	2 1 2 7	3 0 1 8
Prob	$\frac{8\times7\times6}{12\times11\times10}$ $= 0.255$	$0.255 \times \frac{3\times4}{1\times6}$ $= 0.510$	$0.510 \times \frac{2\times3}{2\times7}$ $= 0.218$	$0.218 \times \frac{1\times2}{3\times8}$ $= 0.018$
(pt #)		(15)	(1, 14a)	
P-Value	1.0	0.745	0.236	0.018

*1-sided, guided by H_{alt}:

π of +ve responses with Active $>$ π of +ve responses with Placebo.

Patient number 18:

	Response		Total
	+	-	
Active Injection	2	1	3
Placebo Injection	4	5	9
	6	6	

All possible tables with a total of 6 +ve responses

	0 3 6 3	1 2 5 4	2 1 4 5	3 0 3 6
Prob	$\frac{6\times5\times4}{12\times11\times10}$ $= 0.091$	$0.091 \times \frac{3\times6}{1\times4}$ $= 0.409$	$0.409 \times \frac{2\times5}{2\times5}$ $= 0.409$	$0.409 \times \frac{1\times4}{3\times6}$ $= 0.091$
(pt #)		(17)	(18)	
P-Value	1.0	0.909	0.500	0.091
(1-sided, as above)				

In the Table, 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 "sided-ness" of the H_{alt} after observing the data!!!

And they used different H_{alt} for different patients!!!

M**essage**: Tail areas for this test are tricky: it is best to lay out all the tables, so that one is clear which tables are being included in which tail!

6.2 Fisher's Exact Test and Rhinoceroses

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

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 km^2 area of the northern Namib Desert and on 60 horned animals in the 22,000 km^2 Etosha National Park. On the basis of over 200 witnessed interactions between horned rhinos and spotted hyenas (Crocura crocura) 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 l 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	B	C	total*
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	3	3	

A vs C

	A	C	A	C	A	C	A	C	total*
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	4	3	
Prob	$\frac{1}{35}$		$\frac{12}{35}$		$\frac{18}{35}$		$\frac{4}{35}$		

“Data and conclusions were premature.”

Agree?