

Course EPIB-634: Survival Analysis & Related Topics [Winter 2007]

Assignment 1

course web pages: www.epi.mcgill.ca/hanley/cxxx/ unless otherwise specified

(username: c6xx ; password: 8 letters, H***J*## both case-sensitive)

- 1 Which is the *single biggest* flaw in the analysis of the scouting injuries [page 3]. List two others that might on their own might be major -- but not nearly as large as the distortion produced by *the big one* !
 - 2 Refer to the first row of Table 1 in the Ayas et al. article "Extended Work Duration and the Risk of Self-reported Percutaneous Injuries in Interns" in JAMA on Sept 6 of 2006. (i) Manually calculate the rate per Intern-Month and the 95% CI (ii) Re-express these using Intern-Year as the unit of experience. (iii) Repeat (i) and (ii) using software such as the EpiTools package for R [see <http://www.epitools.net>], or Stata [help epitab, and iri 498 1000000 17003 1000000], or the SAS GENMOD regression program [in c634 Resources]. (iv) Repeat steps (i) to (iii) for the data from the Psychiatry residency, using a method appropriate to the situation [Rothman, 2002, page 127 says 'such situations are the exception rather than the rule.']

You might find "Exact confidence limits on a Poisson parameter: Excel worksheet" (in Resources) helpful for visualizing Poisson distribution, and for exact CI's and p value calculations when the count is too small to rely on the Gaussian approximation (for p-values, you can also interpolate using the table on page 17; or use the exact Poisson function in Excel
 - 3 (i) Calculate a 95% CI for the SIR and test (at $\alpha = 0.05$ 2-sided) $H_0: SIR=1$ for the Alberta Sour Gas Study [p. 4]. Restrict attention to the 33 vs. 36.3 [Index Area 1970 Cohort Females vs. (1) Southern Alberta excl. Calgary, Lethbridge, & Medicine Hat (RP1)]. Describe your procedures/ steps.

Can think of the 'expected number' E as that for all of Alberta, but scaled down to the size of the index area. Because the number of cases for all of Alberta is quite large, it remains stable when we scale it down to E ; Thus, we say that (at least relative to the observed number O in the index area) the scaled down number E has no statistical variation i.e it is treated as a 'constant' in the SIR -- only the numerator O is a random variable.

(ii) Carry out the same tasks, but imagine the concerned area or cohort was much smaller, and that 3 cases were observed where 0.45 were expected. Again, describe your procedures/ steps.
 - 4 Refer to rows 2 and 3 of Table 3 to the Ayas et al article. (i) Manually calculate ORs and 95% CIs, and repeat by computer software. (ii) Explain why your answers do not match those reported (hint: see the paragraph beginning "To assess the relationships..." in the last column of page 1057 of the article. (iii) exactly what (and how many) numbers would you need to carry out *their* analysis for row 3 (injuries in ICU). Answer in the form of a 1-paragraph request to the authors asking for these specific numbers (but do not e-mail the authors! JH has in fact obtained these numbers from Dr Ayas, and they will form the basis for some of next week's homework). (iv) Is OR the correct term for the ratio being estimated here?
 - 5 Refer to the data from John Snow's study, given on bottom of column 3 of page 1 of attached handout for Sept. 05 lecture for Med2 [taken from med2 website, reachable from link at top of 634 website: username med2, password: same as for the cxxx epidemiology courses]. (i) Calculate a 95% CI to accompany the rate ratio of 13.3. Do the same for the ratio estimates based on the denominator series of 100 and 1000 (first column, page 2... [in practice, you would not observe the quasi-denominators shown there, but rather these expected numbers \pm some sampling variation]. (ii) Why are the CI's based on the 100 or 1000 *wider* than the one based on the actual "return which was made to Parliament"?
 - 6 Refer to pages 2 & 3 of the Med2 handout of Nov. 11 [attached]. In dealing with CI's for ratios, it used the fact that for **log-based CI's** (instead of the usual \pm a margin of error for 'regular' statistics) for **ratios**, one can calculate a "multiplied-by/divided-by" factor in order to arrive at the upper/lower limits. (i) Hand-calculate the CI's for the ratio of 13.1 on page 2, and the 1.44 ratio on page 3*, by your usual manual way, and compare them with the answers from the "multiplied-by/divided-by" method shown (ii) Which method do you prefer? (if you have software that does it for you, this is merely a conceptual issue!)
- { * the full article "A population-based study of measles, mumps, and rubella vaccination and autism" can be found under **Nov. 11 lecture in med2**, reachable from link at top of 634 }

/over

7 [OPTIONAL] The following questions concern the article A CONTROLLED TRIAL OF A HUMAN PAPILLOMAVIRUS TYPE 16 VACCINE and are adapted from question 2 in the December 2002 exam in course 626. The abstract is given below; if needed, the full article can be reached via the link on the course 626 webpage. The conditional approach in this question is *similar to that behind the worked e.g. on page 29 of JH's Notes on Poisson Distribution*

- a. The study employed a fixed-numbers of events design" (1st sentence Statistical Analysis section). Why this design rather than a "fixed number of woman-years-of-follow-up" design?
- b. With I denoting incidence, v denoting the vaccinated and u the unvaccinated, Efficacy (E) is defined here as a percentage

$$E = 100 \times (I_u - I_v) / I_u = 100 \times (1 - I_v / I_u)$$

Consider a very large R.C.T. (so random variation is not an issue), with 1/2 receiving the vaccine and 1/2 the placebo, and concentrate on the **total number of cases** (of persistent infection).

What is the relation between the proportion (P) of these cases that would be in the vaccinated group (i.e. what fraction of cases would be 'exposed' cases) and the vaccine efficacy E? To answer, calculate for every 1 case in the unvaccinated, how many cases there would be in the vaccinated; then express the #v as a proportion of (#u + #v).

E (%)	0	25	50	75	80	90
# u cases	1	1	1	1	1	1
# v cases	_____	_____	_____	_____	_____	_____
P=proportion of cases that received v	_____	_____	_____	_____	_____	_____

- c. Suppose that in the actual (finite) study, subject as it was to random variations, the authors had analyzed the data when the total number of cases was 31, i.e. when the observed proportion of cases that had been vaccinated was $p=0/31$ i.e., when the point estimate for P was $p=0.0$. This point estimate translates into an 'exact' 95% 2-sided [binomial-based] CI for P of 0.0 to 0.11.

From this CI, and interpolation in the table you constructed in part b, find 'exact' 95% limits for E.

8 The large-sample methods for obtaining a CI for a rate ratio are accurate when there are enough events in each of the compared categories. But in Q7 above, and in the "Women are Safer Pilots" example on page 3, the small number of events in one of the categories renders large-sample methods inaccurate or even impossible. In such situations, the *conditional* approach, in which one bases the inference on the distribution of the number of events in one category, *conditional* on the *sum* of the numbers of events in the two categories, is a way around this problem (we use a similar *conditioning* strategy when dealing with Fisher's exact test).

Compare the rate of accidents in women relative to men pilots (i.e. the rate *ratio*) (i) Assume that on average, the women pilots fly just as many hours as the men pilots, and that all other relevant factors are equal [although they probably are not!]. Based on the information given, use software to calculate an exact CI for the rate ratio (ii) Repeat, but now assume that on average the women pilots fly half as many hours as the men. (iii) In your own words, and using pages p 29 of Poisson notes, try to describe the basis for the exact method. [JH will use your answers to judge how clear or muddled his description is!]

9 [OPTIONAL, and looking ahead to the link between risk, or cumulative incidence, and rate] Using the reported PI rate for OB/GYN (Table 1 of Ayas et al), what is the probability that an average-risk ob/gyn resident would have no, at least one, PI by the end of (i) 1 month (ii) 12 months of experience? i.e. what is the probability of 'surviving' these lengths of time without a PI?

Is Scouting Safe?

Over the past year, leaders have been showing a growing commitment to provide each member a safe and enjoyable Scouting experience. In support of efforts in the field, we conducted a study to establish baseline data on scouting accident and injury trends so that we can make informed decisions about activity precautions or the need for higher safety standards. This column highlights the findings. The first question we asked ourselves was, "Is Scouting a safe program for members?"

Statistics Canada, Health Division, told us that 11 out of every 1,000 males aged 5-19 are hospitalized for at least one night a year. When we compared similar information taken from Scouting accident forms, we found our members are hospitalized at a rate of only one per thousand a year. Given that we run active programs and heavily use the outdoors, Scouting falls far below the average rate for daily living risk to males in this age group.

Having established this, let's look at the main kinds of accidents and injuries that do happen to Scouting members. Our study identified the types of injuries that happened during the course of a normal Scouting year, excluding summer camps (Chart A). It also recorded the types of activities associated with the injuries (Chart B). When we examine the two sets of information, we begin to see some relationships...

The (Scouting) Leader (magazine) June/July 1991

A CONTROLLED TRIAL OF A HUMAN PAPILLOMAVIRUS TYPE 16 VACCINE

Background Approximately 20 percent of adults become infected with human papillomavirus type 16 (HPV-16). Although most infections are benign, some progress to anogenital cancer. A vaccine that reduces the incidence of HPV-16 infection may provide important public health benefits.

Methods In this double-blind study, we randomly assigned 2392 young women (defined as females 16 to 23 years of age) to receive three doses of placebo or HPV-16 virus-like –particle vaccine (40 µg per dose), at day 0, month 2, and month 6. Genital samples to test for HPV-16 DNA were obtained at enrollment, one month after the third vaccination, and every six months thereafter. Women were referred for colposcopy according to a protocol. Biopsy tissue was evaluated for cervical intraepithelial neoplasia and analyzed for HPV-16 DNA with use of the polymerase chain reaction. The primary end point was persistent 16 infection, defined as the detection of HPV-16 DNA in samples obtained at two or more visits. The primary analysis was limited to women who were negative for HPV-16 DNA and HPV-16 antibodies at enrollment and HPV-16 DNA at month 7.

Results The women were followed for a median of 17.4 months after completing the vaccination regimen. The incidence of persistent HPV-16 infection was 3.8 per 100 woman-years at risk in the placebo group and 0 per 100 woman-years at risk in the vaccine group (100 percent efficacy; 95 percent confidence interval, 90 to 100; $P < 0.001$). All nine cases of HPV-16–related cervical intraepithelial neoplasia occurred among the placebo recipients.

Conclusions Administration of this HPV-16 vaccine the incidence of both HPV-16 infection and HPV-16–related cervical intraepithelial neoplasia. Immunizing HPV-16–negative women may eventually reduce the incidence of cervical cancer. (N Engl J Med 2002;347:1645-51.)

WOMEN ARE SAFER PILOTS: STUDY

LONDON- Initial results of a study by Britain's Civil Aviation Authority shows that women behind the controls of a plane might be safer than men. The study shows that male pilots in general aviation are more likely to have accidents than female pilots. Only 6 per cent of Britain's general aviation pilots are women. According to the aviation magazine Flight International, there have been 138 fatal accidents in general aviation in the last 10 years, and only two involved women - less than 1.5 per cent of the total.

Woman News, page F1 The Montreal Gazette, August 21st, 1995

HURRICANES

USA TODAY, August 1995*. The number of hurricanes with winds greater than 110 mph has declined since the 1950s. Numbers of major hurricanes to hit the USA each decade:

1900s: 6; 1910s: 8; '20s: 5; '30s: 8; '40s: 8; '50s: 9; '60s: 6; '70s: 4; '80s: 6; '90s: 2

[* The 1995 'season' is excluded; i.e., the 1990s data are from the 5 seasons 1990, 1991, 1992, 1993 and 1994]

Table 15.1.4/1 from Alberta Sour Gas Study (Spitzer et al. 1985)

Standardized Incidence Ratios (SIR) excluding non-melanotic skin cancer for the index area 1970 cohort relative to three reference populations

Population	Observed	Expected (1)	Expected (2)	Expected (3)
Index Area 1970 Cohort <i>Males</i>	45	49.6	49.6	52.7
Index Area 1970 Cohort <i>Females</i>	33	36.3	34.2	35.1
Index Area 1970 Cohort Total	78	85.9	83.8	87.8
SIR		0.91	0.93	0.89
95% CI		(0.67 to 1.26)	0.68 to 1.29	0.65 to 1.23

(1) Southern Alberta excl. Calgary, Lethbridge, & Medicine Hat (RP1)

(2) Census Division 6 excl. Southern Alberta excl. Calgary (RP2) - contains DCCI

(3) Census Division 2 excl. Lethbridge (RP3) - contains SR

Lecture September 05

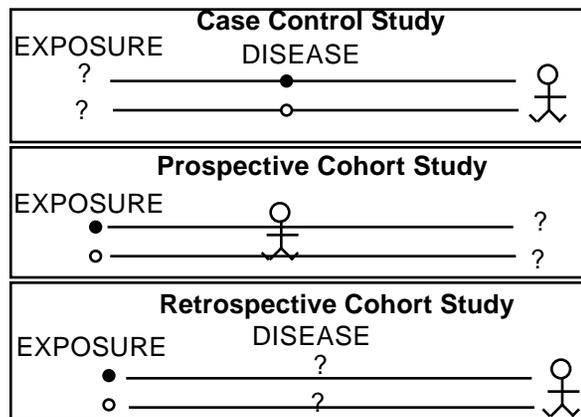
Case-Control Studies [Fletcher Ch10]

Recall [excerpt from Rothman & Greenland] .. there are two primary types of non-experimental studies in epidemiology.

The first, the **cohort study** (also called *the follow-up study* or *incidence study*), is a direct analogue of the experiment; different exposure groups are compared, but (as in Snow's study) the investigator does not assign the exposure.

The other, the incident case-control study, or simply the **case-control study**, employs an extra step of sampling according to the outcome of individuals in the population. This extra sampling step can make a case-control study much more efficient than a cohort study of the entire population, but it introduces a number of subtleties and avenues for bias that are absent in typical cohort studies.

*** e.g. Hennekens and Buring Fig 2-3



● Present basis on which groups are selected at beginning of study
 ○ Absent
 ? to be determined investigator at beginning of study

Case-control studies are best understood by defining a source population, which represents a hypothetical study population in which a cohort study might have been conducted. If a cohort study were undertaken, the primary tasks would be to identify the exposed and unexposed denominator experience, measured in person-time units of experience or as the number of people in each study cohort, and then to identify the number of cases occurring in each person-time category or study cohort. In a case-control study, the cases are identified and their exposure status is determined just as in a cohort study, but denominators from which rates could be calculated are not measured. Instead, a control group of study subjects is sampled from the entire source population that gives rise to the cases.

The purpose of the control group is to determine the relative (as opposed to absolute) size of the exposed and unexposed denominators within the source population. From the relative size of the denominators, the relative size of the incidence rates (or incidence proportions, depending on the nature of the data) can be estimated. **Thus, case-control studies yield estimates of relative effect measures.** Because the control group is used to estimate the distribution of exposure in the source population,

In sum, case-control studies of incident cases differ from cohort studies according to how subjects are initially selected. A cohort study identifies and follows a population or populations to observe disease experience; a case-control study involves an additional step of selecting cases and controls from this population. [end of excerpt]

NOTE[JH] The statistical precision of the ratio measure of risk is largely a function of the number of cases. The same amount of person time is needed to generate a given no. of cases in a cohort study as in a case-control study. The latter's efficiency derives from the reduced amount of data-gathering, and the investigator's time-scale -- IF the exposure of past cases and "non-cases" can be accurately established after the fact.

The essential difference can be illustrated using the data from John Snow's investigation

"According to a return which was made to Parliament, the Southwark and Vauxhall Company supplied 40,046 houses from January 1 to December 31, 1853, and the Lambeth Company supplied 26,107 houses during the same period; "

So, the **denominators** were...

No. of Houses with...	
Water	
Impure	Pure
40 046	26 107

286 fatal attacks of cholera took place, in the first four weeks of the epidemic, in houses supplied by the former company, and only 14 in houses supplied by the latter

No. of CASES (numerators) in houses with...

Water	
Impure	Pure
286	14

Attack rates in houses with...

Water			
Impure	Pure	Ratio	Difference
$\frac{286}{40046}$	$\frac{14}{26107}$		
71.4 / 10K	5.4 / 10K	13.3	66 / 10K

This is the cohort approach -- start with denominators of known sizes and then determine the numerators.

But what if the sizes of the two denominators were not readily available (but the numerators were) ???. it would be a lot of leg work to determine the water source of each of 40046 + 26107 = 66153 houses!

No. of CASES (**numerators**) in houses with...

Water	
Impure	Pure
286	14

If a huge amount of work to determine the sizes of the two denominators, how about we take a sample and estimate their estimate their relative sizes ?

Say we survey **100** houses selected at random; we might find that the **sources** were...

No. (± sampling variation) of **100** **sampled Houses** with...

Water		
Impure	Pure	
61 (±10)	39 (±10)	100

We can take the **61** and **39** as "**quasi-denominators**" and make two "**quasi-rates**"

Quasi-attack rates in houses with...

Water			
Impure	Pure	Ratio	Difference
$\frac{286}{61}$	$\frac{14}{39}$	13.1 (±)	no meaning

Lets say that instead we survey **1000** houses selected at random and that the sources were...

No. (± sampling variation) of **1000** **sampled Houses** with...

Water	
Impure	Pure
605 (±32)	395 (±32)

Quasi-attack rates in houses with...

Water			
Impure	Pure	Ratio	Difference
$\frac{286}{605}$	$\frac{14}{395}$	13.3 (±)	no meaning

Thus the purpose of the 100 (or 1000, or however many are selected, depending on the budget, and the statistical precision required) houses selected at random is to determine the relative (as opposed to absolute) size of the exposed and unexposed denominators within the source population. From the relative size of the denominators, the relative size of the incidence rates (or incidence proportions, depending on the nature of the data) can be estimated.

A good descriptor of these houses selected at random is "the **denominator series**". The cases, already in hand, constitute the "numerator series". [terminology of McGill Prof Miettinen]

To make the calculation of the statistical errors associated with the estimated ratio less complicated, most epidemiologists would exclude the "case houses" from the sampling frame of 66153 houses and would instead sample the "source to be determined" houses from the remainder - i.e. from the "non-case houses". See for example Fletcher et al.'s Figure 10.3, where they write of "non-cases".

Unfortunately, the more common (and older) name for these "non-case" houses is the "control" houses. This creates considerable **confusion** among non-epidemiologists, since we **now have 2 meanings for "control"** ..

1 in an experiment (e.g. clinical trial), those who do not receive the experimental (new) treatment are sometimes referred to as the "controls" ("comparison group" or --if it is the situation -- "unexposed group" is a more informative label) The same applies in a (non-experimental) cohort study (e.g. what should one call the wives of the male resident physicians when their pregnancy outcomes are compared with those of the female resident physicians?)

Notice that Fletcher et al. themselves use confusing terminology -- in describing the characteristics of a cohort study (Table 10.2 3rd row, 1st column) they say "Controls, the

comparison group (i.e. noncases), not selected -- evolve naturally.

2 in a "study that relies on quasi-denominators", (commonly known as a "case-control" study), the "controls" are the denominator series. Their exposure status (or exposure history) is the focus of the inquiry. Even though it is not entirely accurate, it is less confusing to call them "non-cases" than to call them "controls".

"Being epidemiologically correct"... Most epidemiology textbooks still describe case-control studies as "comparing cases with controls". In fact, as the above example [that views the "controls (or non-cases) as a denominator series] shows, **even in a case-control study one compares** (quasi-rates) for the **exposed** with quasi-rates for the **non-exposed** (in the ratio of these quasi-rates, the hidden sampling fraction cancels out in the arithmetic)

This last point about the sampling fraction is very important: the "controls" [i.e., the "non-case" or "the denominator series"] must be selected without regard to their exposure.. see page 1 re "this cardinal requirement"

Other simple e.g.'s of denominator issue:

"Pour battre Patrick Roy, mieux vaut lancer bas" (JH course 626)

WOMEN ARE SAFER PILOTS: newspaper article (JH course 626)

Could we use a case-control approach to the Study of Medical students' compliance with simple administrative tasks and success in final examinations?

Most important issues in case-control studies

- appropriate "controls"
 - "nested" case-control approach attractive i.e. explicit source (e.g.. Medicare database of all Saskatchewan residents: tumor registry; database of all prescription drugs dispensed [universal drug coverage])
- selection bias (e.g. naive md's contributed only the "exposed cases" in study of 3rd generation OC's and risk of venous thromboembolic disorders)
- exposure and confounder ascertainment

Introduction to Statistical Inference*

Inference is about Parameters (Populations) or general mechanisms -- or future observations. It is not about data (samples) *per se*, although it uses data from samples. Might think of inference as statements about a universe most of which one did not observe, or has not yet observed .

Two main schools or approaches:

Bayesian [not even mentioned by Fletcher]

- Makes direct statements about parameters and future observations
- Uses previous impressions plus new data to update impressions about parameter(s)

e.g.
 Everyday life
 Medical tests: Pre- and post-test impressions

Frequentist

- Makes statements about observed data (or statistics from data) (used indirectly [but often incorrectly] to assess evidence against certain values of parameter)
- Does not use previous impressions or data outside of current study (meta-analysis is changing this)

e.g.
 • Statistical Quality Control procedures [for Decisions]
 • Sample survey organizations: Confidence intervals
 • Statistical Tests of Hypotheses

Unlike Bayesian inference, there is no quantified pre-test or pre-data "impression"; the ultimate statements are about data, conditional on an assumed null or other hypothesis.

Thus, an explanation of a p-value must start with the conditional "IF the parameter is ... the probability that the data would ..."

Book "Statistical Inference" by Michael W. Oakes is an excellent introduction to this topic and the limitations of frequentist inference.

(Frequentist) Confidence Interval (CI) or Interval Estimate for parameter

Formal definition:

A level $1 - \alpha$ Confidence Interval for a parameter θ is given by two statistics (i.e.. numbers calculated from data)

Upper and Lower

such that when θ is the true value of the parameter,

$$\text{Prob} (\text{Lower} \leq \theta \leq \text{Upper}) = 1 - \alpha$$

α	$1 - \alpha$
0.05	0.95
0.01	0.99

- CI is a **statistic**: a quantity calculated from a sample
- usually use $\alpha = 0.01$ or 0.05 or 0.10 , so that the "level of confidence", $1 - \alpha$, is 99% or 95% or 90%. We will also use " α " ("alpha") for tests of significance (there is a direct correspondence between confidence intervals and tests of significance)
- technically, we should say that we are **using a procedure which is guaranteed to cover the true θ in a fraction $1 - \alpha$ of applications**. If we were not fussy about the semantics, we might say that any particular CI has a $1 - \alpha$ chance of covering θ .
- for a given amount of sample data] the narrower the interval from L to U, the lower the degree of confidence in the interval and vice versa.

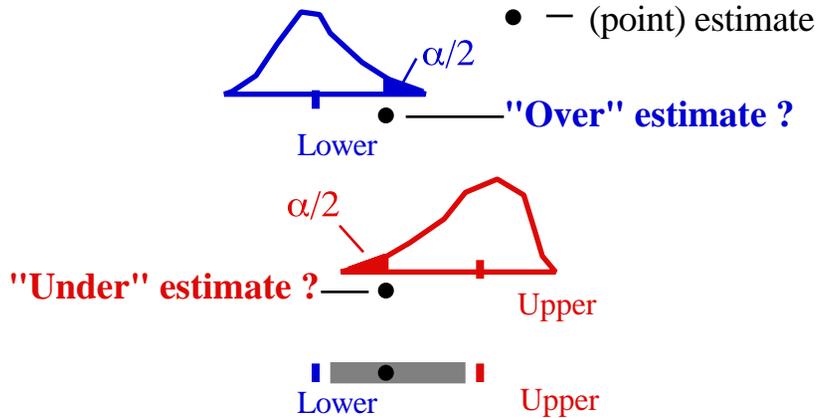
Large-sample CI's, based on Standard Error (SE) of statistic

Many large-sample CI's are of the form ($\hat{\theta}$ denotes 'estimate of')

- i $\hat{\theta} \pm \text{multiple of } SE(\hat{\theta})$ or
- ii inverse fn. of [$f(\hat{\theta}) \pm \text{multiple of } SE(f(\hat{\theta}))$]. where fn. is some function of $\hat{\theta}$ which has close to a Gaussian distribution.
 e.g. $\hat{\theta} = \text{odds or rate ratio}$; $f_n = \ln$ (natural log) ; inv. fn. = exp.

- 'Multiple' based on desired level of 'confidence'
 e.g. 1.645 for 90% confidence, 1.96 for 95% confidence.
- Standard error (SE) is a function of amount of information on which estimate is based (the more the information, the smaller the SE).
- the ' $1.645 \times SE$ ' or ' $1.96 \times SE$ ' called the '**margin of error**'

Method of Constructing a 100(1 - α)% CI (in general):



SE's for "Large Sample" CI's for parameters, and DIFFERENCES thereof [: standard deviation (SD) of individuals]

parameter	estimate	SE[estimate]
	$\hat{\mu}$	$SE[\hat{\mu}]$
mean μ_y	\bar{y}	$\frac{s_y}{\sqrt{n}}$
prop.	p	$\frac{\sqrt{p[1-p]}}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$p_1 - p_2$	$p_1 - p_2$	$\sqrt{\frac{p_1[1-p_1]}{n_1} + \frac{p_2[1-p_2]}{n_2}}$

prop. = proportion

In SE, estimated values substituted for unknown ones (in Greek)

EXAMPLES: See exercise on Birthweights and Adult Heights

"Large Sample" CI for Odds Ratio

parameter data and odds ratio (or rate ratio) estimate
(denominators: full [cohort] or samples ['controls'])

	Exposed(1)	Not(0)
Odds Ratio	$\frac{\#cases}{\#denominator}$	$\frac{\#cases}{\#denominator}$

SE[log odds ratio]

$$= \sqrt{\frac{1}{\#exposed\ cases} + \frac{1}{\#unexposed\ cases} + \frac{1}{\#exposed\ denominator} + \frac{1}{\#unexposed\ denominator}}$$

EXAMPLE (Kim 2002): No. of CASES of nasal polyposis (numerators) among people who live in houses heated by ...

Woodstove ?

Yes (1)	No (0)	Total
45	10	55 (CASE series)

No. of **sampld** (same age-sex) **people** who live in houses heated by...

Woodstove ?

Yes (1)	No (0)	Total
14	41	55 ('denominator' series, 'CONTROLS')

Quasi-rates in people who live in houses heated with... Woodstove

Yes (1)	No (0)	ratio	MARGIN OF ERROR i.e. multiplier and divisor to be applied to point estimate (i.e. to observed ratio)	95% CI
41	10	13.1	$\exp[1.96 \times \sqrt{\frac{1}{45} + \frac{1}{10} + \frac{1}{14} + \frac{1}{41}}]$	13.1 ÷ 2.5 to 13.1 × 2.5
14	41		= 2.5	5.2 to 32.8

NOTE: If denominator *much larger* than # cases (as in cohort study), SE of log odds ratio dominated by # exposed **cases** and # unexposed **cases**. (Control:case ratio of 4 => SE = $\sqrt{1/1 + 1/4} = 1.12 \times \sqrt{1/1 + 1/4}$).

"Large Sample" CI for Rate Ratio AUTISM & MMR vaccinations

No. of CASES of autism (numerators) among children who did / did not receive MMR vaccination ... Danish Cohort Study, NEJM Nov 7, 2002

Vaccinated		
Yes (1)	No (0)	
263	53	316 (CASES)

No. of children-years (c-y) of follow-up [contributed by 0.54 m children]

Vaccinated		
Yes (1)	No (0)	
1.65m	0.48m	2.13m children-years (c-y) (DENOMINATORS)

CRUDE Rates ...

Vaccinated				
Yes (1)	No (0)	rate ratio	margin of error i.e. multiplier and divisor to be applied to point estimate (i.e. to observed rate ratio)	95% CI*
$\frac{263}{1.65m}$	$\frac{53}{0.48m}$	1.44 *	$\exp[1.96 \times \sqrt{\frac{1}{263} + \frac{1}{53}}]$ = 1.34 †	1.44 ÷ 1.34 to 1.44 × 1.34 1.07 to 1.93

† SE of log rate ratio determined by numbers of cases.

* Note the big difference between the crude (1.44) and adjusted ratio (reason why discussed next). For this reason, I am calculating the CI around the crude ratio in this example simply for didactic purposes. The adjusted ratio was 0.92, 95% CI 0.68 to 1.24 (i.e., 0.92 × ÷ 1.35)

Rate ratio: "crude"= 1.44; Adjusted (cf. article) = 0.92. WHY ?

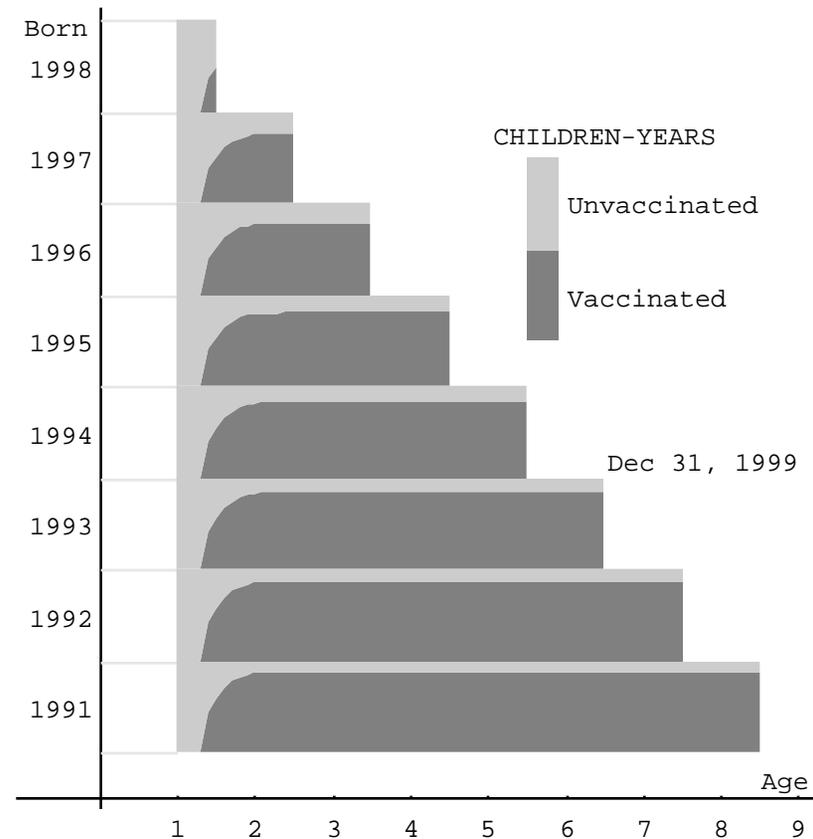
"We calculated the relative risk with adjustment for age, calendar period, sex, birthweight, gestational age, mother's education, and socio-economic status"

(p 1479, 2nd column, 6 lines from end)

"Except for age, none of these possible confounders changed the estimates. The confounding by age was a function of the time available for follow-up, since much of the follow-up for the unvaccinated group involved young children, in whom autism is often unnoticed"

(p 1481, 2nd column, end 1st paragraph)

Note[JH]: cf. footnote regarding missing gestational ages, Table 1. (Schematic to help visualize the confounding .. constructed by JH)



Pair-matched Case-Control Study

(e.g. 1st article used in small group session 1, and again in session 2)

The formula on page 1 is for an *unmatched* analysis. For a *matched-pair* case control analysis, with 1 denoting exposed and 0 unexposed,

$$\text{odds ratio (i.e., point estimate)} = \frac{\# \{ \text{case}_1 ; \text{control}_0 \} \text{ pairs}}{\# \{ \text{case}_0 ; \text{control}_1 \} \text{ pairs}}$$

SE[log odds ratio]

$$= \sqrt{\frac{1}{\# \{ \text{case}_1 ; \text{control}_0 \} \text{ pairs}} + \frac{1}{\# \{ \text{case}_0 ; \text{control}_1 \} \text{ pairs}}}$$

The (many) ways to (in)correctly describe a CI

Below are my annotated answers to some graduate students' interpretations of a CI

Question: A New York Times poll on women's issues interviewed 1025 women and 472 men randomly selected from the United States excluding Alaska and Hawaii. The poll found that 47% of the women said they do not get enough time for themselves.

- (a) The poll announced a margin of error of ± 3 percentage points for 95% confidence in conclusions about women. Explain to someone who knows no statistics why we can't just say that 47% of all adult women do not get enough time for themselves.
- (b) **Then explain clearly what "95% confidence" means.**
- (c) The margin of error for results concerning men was ± 4 percentage points. Why is this larger than the margin of error for women?

- | | | |
|---|--|---|
| 1 | True value will be between 43 & 50% in 95% of repeated samples of same size. | • No. Estimate will be between $\mu - \text{margin}$ & $\mu + \text{margin}$ in 95% of applications. |
| 2 | Pollsters say their survey method has 95% chance of producing a range of percentages that includes . | • Good. Emphasize average <u>performance</u> in repeated <u>applications</u> of method. |
| 3 | If this same poll were repeated many times, then 95 of every 100 such polls would give a range that included 47%. | • No! . See 1. |
| 4 | You're pretty sure that the true percentage is within 3% of 47%. "95% confidence" means that 95% of the time, a random poll of this size will produce results within 3% of . | • Bayesians would object (and rightly so!) to this use of the "true parameter" as the <u>subject</u> of the sentence. They would insist you use the statistic as the subject of the sentence and the parameter as object. |

- | | | |
|----|--|--|
| 5 | If took 100 different samples, in 95% of cases, the sample proportion will be between 44% and 50%. | • NO! The sample proportion will be between truth - 3% & truth + 3% in 95% of them. |
| 6 | With this one sample taken, we are sure 95 times out of 100 that 41-53% of the women surveyed do not get enough time for themselves. | • NO. 95/100 times the estimate will be within 3% of , i.e., estimate will be in interval - margin to + margin. Method used gives correct results 95% of time. |
| 7 | In 95 of 100 comparable polls, expect 44 - 50% of women will give the same answer.

Given a parameter, we are 95% sure that the mean of this parameter falls in a certain interval. | • NO. Same answer? as what?

Not given a parameter (ever) . If we were, wouldn't need this course!
Mean of a parameter makes no sense in frequentist inference. |
| 8 | "using the poll procedure in which the CI or rather the true % is within +/- 3, you cover the true percentage 95% of times it is applied. | • A bit muddled ... but "correct in 95% of applications" is accurate. |
| 9 | Confident that a poll (such) as this one would have measured correctly that the true proportion lies between in 95% . | • ??? [I have trouble parsing this!]
In 95% of applications/uses, polls like these come within $\pm 3\%$ of truth. |
| 10 | 95% chance that the info is correct for between 44 and 50% of women. | • ??? 95% confidence in the procedure that produced the interval 44-50 |
| 11 | 95% confidence -> 95% of time the proportion given is the good proportion (if we interviewed other groups). | • "Correct in 95% of applications"
Good to connect the 95% with the long run, not specifically with this one estimate. Always ask yourself: what do I mean by "95% of <u>the time</u> " ? If you substitute "applications" for "time", it becomes clearer. |
| 12 | It means that 47% give or take 3% is an accurate estimate of the population mean 19 times out of 20 such samplings.
"This result is trustworthy 19 times out of 20"

"this poll" : see COMMENT below<---- | • ??? 95% of applications of CI give correct answer. How can the <u>same</u> interval $47\% \pm 3$ be accurate in 19 but not in the other 1?

• ??? "this" result: Cf. the distinction between " <u>my</u> operation is successful 19 times out of 20 ..." and "operations <u>like the one to be done on me</u> are successful 19 times out of 20" |

COMMENT: Polling companies who say "polls of this size are accurate to within so many percentage points 19 times out of 20" are being statistically correct -- they emphasize the procedure rather than what has happened in this specific instance. Polling companies (or reporters) who say "this poll is accurate .. 19 times out of 20" are talking statistical nonsense -- this specific poll is either "right" or "wrong"!. On average 19 polls out of 20 are "correct" . But **this poll cannot be right on average 19 times out of 20!**

Even more ways to (in)correctly describe a CI

The Gallup Poll asked 1571 adults what they considered to be the most serious problem facing the nation's public schools; 30% said drugs. This sample percent is an estimate of the percent of all adults who think that drugs are the schools' most serious problem. The news article reporting the poll result adds, "The poll has a margin of error -- the measure of its statistical accuracy -- of three percentage points in either direction; aside from this imprecision inherent in using a sample to represent the whole, such practical factors as the wording of questions can affect how closely a poll reflects the opinion of the public in general" (The New York Times, August 31, 1987). The Gallup Poll uses a complex multistage sample design, but the sample percent has approximately a normal distribution. Moreover, it is standard practice to announce the margin of error for a 95% confidence interval unless a different confidence level is stated.

- a The announced poll result was 30%±3%. Can we be certain that the true population percent falls in this interval? -->
- b Explain to someone who knows no statistics what the announced result 30%±3% means. ANNOTATED ANSWERS next column... -->
- c Does the announced margin of error include errors due to practical problems such as undercoverage and nonresponse? ANSWER: NO!

Meta-analysis

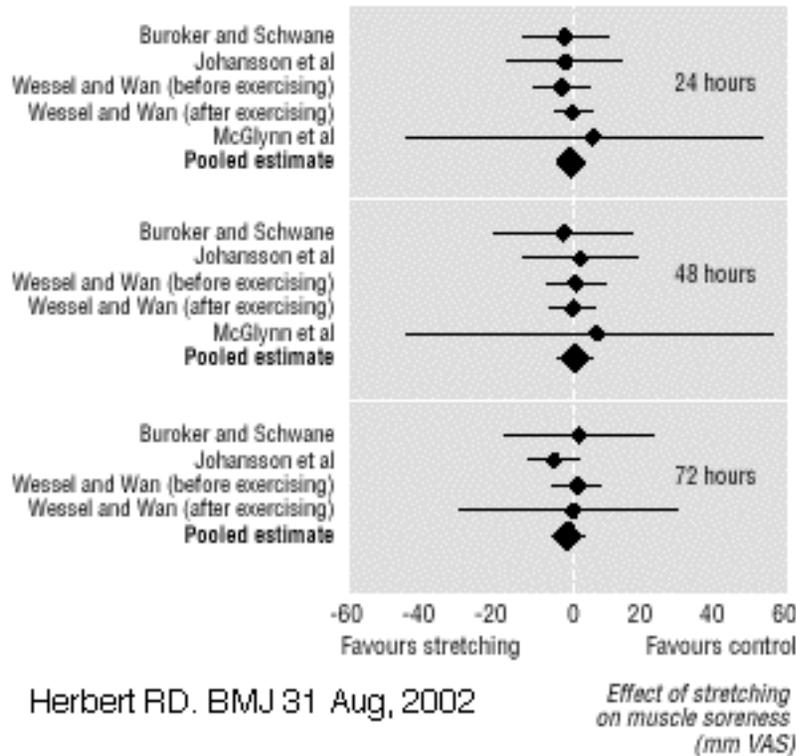


Fig 1 Effects of stretching on delayed onset muscle soreness at 24 hours, 48 hours, and 72 hours after exercise. (VAS=visual analogue scale) 15-17 19

- 1 This means that the population result will be between 27% and 33% 19/20 times.
- 2 95% of the time the actual truth will be between 30 ± 3% and 5% it will be false.
- 3 If this poll were repeated very many times, then 95 of 100 intervals would include 30% .
- 4 Interval of true values ranges b/w 27% + 33%.
- 5 Confident that in repeated samples estimate would fall in this range 95/100 times.
- 6 95% of intervals will contain true parameter value and 5% will not. Cannot know whether result of applying a CI to a particular set of data is correct.
- 7 In 1/20 times, the question will yield answers that do not fall into this interval.
- 8 This type of poll will give an estimate of 27 to 33% 19 times out of 20 times.
- 9 5% risk that μ is not in this interval.
- 10 95 / 100 times if do the calculations, result 27-33% would appear.
- 11 95% prob computed interval will cover parameter.
- 12 The true popl'n mean will fall within the interval 27-33 in 95% of samples drawn.

- **NO! Population result is wherever it is and it doesn't move.** Think of it as if it were the speed of light.
- It either is or it isn't ... the truth doesn't vary over samplings.
- **NO.** 95% of polls give answer within 3% of truth, NOT within 3% of the mean in this sample.
- **???** There is only one true value. AND, it isn't 'going' or 'ranging' or 'moving' anywhere!
- **NO.** Estimate falls within 3% of in 95% of applications
- **GOOD.** Say "Cannot know whether CI derived from a particular set of data is correct." **Know about behaviour of procedure!** If not from Mars, (i.e. if you use past info) might be able to bet more intelligently on whether it does or not.
- **No.** In 5% of applications, estimate will be more than 3% away from true answer. See 1,2,3 above.
- **NO.** Won't give 27 ± 3 19/20 times. Estimate will be within ± 3 of truth in 19/20 applications
- **???** If an after the fact statement, somewhat inaccurate.
- **No it wouldn't.** See 1,2,3,7.
- Accurate if viewed as a prediction.
- **NO.** True popl'n mean will not "fall" anywhere. It's a fixed, unknowable constant. Estimates may fall around it.

1200 are hardly representative of 80 million homes /220 million people!!

The Nielsen system for TV ratings in U.S.A.
(Excerpt from article on "Pollsters" from an airline magazine)

"...Nielsen uses a device that, at one minute intervals, checks to see if the TV set is on or off and to which channel it is tuned. That information is periodically retrieved via a special telephone line and fed into the Nielsen computer center in Dunedin, Florida.

With these two samplings, Nielsen can provide a statistical estimate of the number of homes tuned in to a given program. A rating of 20, for instance, means that 20 percent, or 16 million of the 80 million households, were tuned in.

To answer the criticism that 1,200 or 1,500 are hardly representative of 80 million homes or 220 million people, Nielsen offers this analogy:

Mix together 70,000 white beans and 30,000 red beans and then scoop out a sample of 1000. the mathematical odds are that the number of red beans will be between 270 and 330 or 27 to 33 percent of the sample, which translates to a "rating" of 30, plus or minus three, with a 20-to-1 assurance of statistical reliability. The basic statistical law wouldn't change even if the sampling came from 80 million beans rather than just 100,000." ...

Why, if the U.S. has a 10 times bigger population than Canada, do pollsters use the same size samples of approximately 1,000 in both countries?

Answer : it depends on **WHAT IS IT THAT IS BEING ESTIMATED**. With $n=1,000$, the SE or uncertainty of an estimated **PROPORTION** 0.30 is indeed 0.03 or 3 percentage points. However, if interested in the **NUMBER** of households tuned in to a given program, the best estimate is $0.3N$, where N is the number of units in the population ($N=80$ million in the U.S. or $N=8$ million in Canada). The uncertainty in the 'blown up' estimate of the **TOTAL NUMBER** tuned in is blown up accordingly, so that e.g. the estimated **NUMBER** of households is

U.S.A. $80,000,000[0.3 \pm 0.03] = 24,000,000 \pm 2,400,000$
Canada. $8,000,000[0.3 \pm 0.03] = 2,400,000 \pm 240,000$

2.4 million is a 10 times bigger absolute uncertainty than 240,000. Our intuition about needing a bigger sample for a bigger universe probably stems from absolute errors rather than relative ones (which in our case remain at 0.03 in 0.3 or 240,000 in 2.4 million or 2.4 million in 24 million i.e. at 10% irrespective of the size of the universe.

It may help to think of why we do not take bigger blood samples from bigger persons: the reason is that we are usually interested in concentrations rather than in absolute amounts and that concentrations are like proportions.

The "Margin of Error blurb" introduced (legislated) in the mid 1980's

Montreal Gazette August 8, 1981

NUMBER OF SMOKERS RISES BY FOUR POINTS: GALLUP POLL

Compared with a year ago, there appears to be an increase in the number of Canadians who smoked cigarettes in the past week - up from 41% in 1980 to 45% today. The question asked over the past few years was: "**Have you yourself smoked any cigarettes in the past week?**" Here is the national trend:

Year	'74	'75	'76	'77	'78	'79	'80	'81
Smoked cigarettes in past week (%)	52	47	??	45	47	44	<u>41</u>	<u>45</u>

Today's results are based on 1,054 personal in-home interviews with adults, 18 years and over, conducted in June.

The Gazette, Montreal, Thursday, June 27, 1985

39% OF CANADIANS SMOKED IN PAST WEEK: GALLUP POLL

Almost two in every five Canadian adults (39 per cent) smoked at least one cigarette in the past week - down significantly from the 47 percent who reported this 10 years ago, but at the same level found a year ago. Here is the question asked fairly regularly over the past decade: "**Have you yourself smoked any cigarettes in the past week?**" The national trend shows:

Year	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85
Smoked cigarettes in past week (%)	47	??	45	47	44	41	45	42	41	39	39

^^ Smoked regularly or occasionally? [JH: larger n won't reduce 'non-sampling' variation]

Results are based on 1,047 personal, in-home interviews with adults, 18 years and over, conducted between May 9 and 11. A sample of this size is accurate within a 4-percentage-point margin, 19 in 20 times.

La Presse, Montréal, 1993

95%CI? IC? ... Comment dit on... ?

L'Institut Gallup a demandé récemment à un échantillon représentatif de la population canadienne d'évaluer la manière dont le gouvernement fédéral faisait face à divers problèmes économiques et général. Pour 59 pour cent des répondants, les libéraux n'accomplissent pas un travail efficace dans ce domaine, tandis que 30 pour cent se déclarent de l'avis contraire et que onze pour cent ne formulent aucune opinion.

La même question a été posée par Gallup à 16 reprises entre 1973 et 1990, et ne n'est qu'une seule fois, en 1973, que la proportion des Canadiens qui se disaient insatisfaits de la façon dont le gouvernement gérait l'économie a été inférieure à 50 pour cent.

Les conclusions du sondage se fondent sur 1009 interviews effectuées entre le 2 et le 9 mai 1994 auprès de Canadiens âgés de 18 ans et plus. Un échantillon de cette ampleur donne des résultats exacts à 3,1 p.c., près dans 19 cas sur 20. La marge d'erreur est plus forte pour les régions, par suite de l'importance moindre de l'échantillonnage; par exemple, les 272 interviews effectuées au Québec ont engendré une marge d'erreur de 6 p.c. dans 19 cas sur 20.