1. Comparison of 2 Poisson Counts
2. Comparison of 2 Rates
3. Power/Precision: sample sizes
4. Extra-Poisson Variation
5. CI and test of $\lambda_1 - \lambda_2$, and λ_1/λ_2, by Poisson regression

1 Preamble

These notes are based in part on JH's courses for epidemiologists. They are less about the 'wiring' of the car and more about how to drive it. But they contain some practical material not covered in the 'bite-sized' C&H chapters. And, like the one for a single Poisson random variable, they give more (but less technical) detail on the distributions used. JH plans to eventually fully integrate this material with the (newer) notes and exercises on C&H chapter 13.

2 Comparison of two counts:

- Note from JH: [1]

  A & B also use the letter μ for the mean (or expected) number of events in the amount of experience actually studied (e.g., PT if it is an amount of *P*opulation-*T*ime. Say y (or c, for 'cases') is the observed number of events. With this notation, we can use λ and λ as the theoretical and empirical/observed rate or incidence density, i.e., $\lambda = \mu/PT$, $\hat{\lambda} = y/PT$. Some textbooks use λ where A & B , and JH, would use μ.

  The excerpts below (indended) are from A & B; however JH has changed the realizations to y_1 and y_2 rather than the x_1 and x_2 used by them. He did this to emphasize that when we come to Poisson (or other) regression models, the counts will be the dependent ("y") variable.

- From A & B, p 156 (See also Pocock SJ. BMJ. 2006;332(7552):1256-8)

  > Suppose that y_1 is a count which can be assumed to follow a Poisson distribution with mean μ_1. Similarly let y_2 be a count independently following a Poisson distribution with mean μ_2. How might we test the null hypothesis that $\mu_1 = \mu_2$?

[1] From Armitage & Berry, Ch 5.2. van Belle, ch. 6.5 deals only with 1-sample problems.

Note (JH): A & B deal with the *simplest* comparative situation, where the amounts of experience *(denominators) are equal,* i.e. $PT_1 = PT_2$, and so a test of $\lambda_1 = \lambda_2$ is equivalent to a test of $\mu_1 = \mu_2 = \mu$.

> **One approach** would be to use the fact that the variance of $y_1 - y_2$ is $\mu_1 + \mu_2$. The best estimate of $\mu_1 + \mu_2$ on the basis of the available information is $y_1 + y_2$. On the null hypothesis $E(y_1 - y_2) = \mu_1 - \mu_2 = 0$, and $y_1 - y_2$ can be taken to be approximately normally distributed unless μ_1 and μ_2 are very small. Hence,
>
> $$z = \frac{y_1 - y_2}{(y_1 + y_2)^{1/2}}$$
>
> can be taken as approximately a standardized normal deviate.

This is a large-sample method: $z = (\hat{\mu_1} - \hat{\mu_2})/\{Var_0[\hat{\mu_1} - \hat{\mu_2}]\}^{1/2}$, with $Var_0[\hat{\mu_1} - \hat{\mu_2}]$ estimated, under the null, i.e., as $Var_0[\hat{\mu_1} - \hat{\mu_2}] = \hat{\mu} + \hat{\mu} = (1/2)(y_1 + y_2) + (1/2)(y_1 + y_2) = y_1 + y_2$, so that $z = (y_1 - y_2)/\{y_1 + y_2\}^{1/2}$.

> A **second approach** has already been indicated in the test for the comparison of proportions in paired samples (section 4.5). Of the *total* frequency $y_1 + y_2$, a portion y_1 is observed in the *first* sample. Writing $r = y_1$ and $n = y_1 + y_2$ in (4.17) we have
>
> $$z = \frac{y_1 - (1/2)(y_1 + y_2)}{(1/2)(y_1 + y_2)^{1/2}} = \frac{y_1 - y_2}{(y_1 + y_2)^{1/2}}$$
>
> as in the first approach. The two approaches thus lead to exactly the same test procedure.

This is a large-sample approximation to a *conditional* test, based on fact (Casella & Berger, p194, ex. 4.15) that if $y = y_1 + y_2$, where y_1 and y_2 are independent Poisson r.v.'s, then $y_1 \mid y \sim$ Binomial$(y, 0.5)$.

y_1 as a *proportion* of $y = (y_1 + y_2)$, tested against a Binomial$(y, 0.5)$.

A & B only give the large-sample version of this test. It is also possible, (and easy in 2012!) to evaluate the p-value using exact Binomial distribution.

> A **third approach** uses a rather different application of the X^2 test from that described for the 2×2 table in section 4.5, the total frequency of $y_l + y_2$ now being divided into two components rather than four. Corresponding to each observed frequency we can consider the expected frequency, on the null hypothesis, to be $(1/2)(y_1 + y_2) = \bar{y}$, say.

Observed:	y_1	y_2
Expected:	$(1/2)(y_1+y_2)$	$(1/2)(y_1+y_2)$

Applying the usual formula (4.30) for a X^2 statistic, we have

$$X^2 = (y_1-\bar{y})^2/\bar{y} + (y_2-\bar{y})^2/\bar{y} = (y_1-y_2)^2/(y_1+y_2).$$

This is a X^2 statistic from a "2×1" table i.e., 2 samples, finite numerators, infinite person-moments of experience. Note that $X^2 = z^2$.

3 Rate Difference $\lambda_1 - \lambda_2$ and Rate Ratio $\lambda_1 \div \lambda_2$

A & B cover only the case where $PT_1 = PT_2$. We deal with the *more general case* of inference re the difference between 2 *rates* λ_1 and λ_2, i.e., when $PT_1 \neq PT_2$, i.e. where a naive comparison of the two numerators makes no sense.

3.1 Test of $H_0 : \lambda_1 = \lambda_2$, or $\lambda_1/\lambda_2 = 1$

Again, three equivalent ways to test this:

1. $z = (\hat{\lambda}_1 - \hat{\lambda}_2)/\{Var_0[\hat{\lambda}_1 - \hat{\lambda}_2]\}^{1/2}$,
   with $Var_0[\hat{\lambda}_1 - \hat{\lambda}_2]$ estimated, under the null, i.e., as...
   $Var_0[\hat{\lambda}_1 - \hat{\lambda}_2] = \hat{\mu}_1/PT_1^2 + \hat{\mu}_2/PT_2^2$,
   where, under H_0, $\hat{\mu}_1 = \hat{\lambda} \times PT_1$ and $\hat{\mu}_2 = \hat{\lambda} \times PT_2$, and
   $\hat{\lambda} = (y_1+y_2)/(PT_1+PT_2)$. (*rate based on 'pooled' data*)
2. y_1 as a *proportion* of $y = (y_1+y_2)$, tested against a Binomial(y, π), where $\pi = PT_1/(PT_1+PT_2)$. Can use Normal approximation to Binomial, or calculate tail-areas exactly from the (asymmetric) Binomial distribution.
3. X^2 statistic from a "2×1" table i.e., 2 samples, finite numerators, infinite – but unequal – person-moments of experience:
   $X^2 = (y_1 - E[y_1])^2/E[y_1] + (y_2 - E[y_2])^2/E[y_2]$.
   Under the null, $E[y_1] = y \times \pi, \quad E[y_2] = y \times (1-\pi)$.

3.2 CIs for Rate Difference $\lambda_1 - \lambda_2$ and Ratio $\lambda_1 \div \lambda_2$

3.2.1 Large-sample methods

Rate Difference: $\hat{\lambda}_1 - \hat{\lambda}_2 \mp z \times SE[\hat{\lambda}_1 - \hat{\lambda}_2]$,
with $SE = (\hat{Var}[y_1]/PT_1^2 + \hat{Var}[y_2]/PT_2^2)^{1/2} = (y_1/PT_1^2 + y_2/PT_2^2)^{1/2}$.

Rate Ratio: (a) $(\hat{\lambda}_1 \div \hat{\lambda}_2) \ \div / \times \ (\exp[ME])$,
where $ME = z \times SE[\log(\hat{\lambda}_1 \div \hat{\lambda}_2)] = z \times (1/y_1 + 1/y_2)^{1/2}$.
Note that with **log-symmetric CI's**, the familiar $-/+$ is replaced by $\div/\times$: we speak of an 'x-fold' ('x-fois') uncertainty range, from θ_{LOWER} to θ_{UPPER}.
(b) Test-based CI for $\lambda_1 \div \lambda_2$: $(\hat{\lambda}_1 \div \hat{\lambda}_2)$ to power of $(1 \mp z/X)$.

3.2.2 Small-sample methods

Rate Difference:

See references for small-sample inference re differences in proportions. Use PT's as binomial denominators. If $y_i > PT_i$, then express PT in smaller units, e.g., in person-days rather than person-years, so that, numerically, $y_i >> PT_i$.

Rate Ratio $\theta = \lambda_1 \div \lambda_2$:

We 'condition out' the nuisance parameter, so as to focus on $\theta = \lambda_1 \div \lambda_2$. To do so, we again rely on the fact that if y_1 and y_2 are independent Poisson r.v.'s with means (expectations) μ_1 and μ_2, then

$$y_1 \mid (y_1 + y_2 = y) \sim binomial[\ y,\ \Pi = \mu_1/(\mu_1+\mu_2)\ ].$$

Thus the observed proportion $y_1 \div (y_1+y_2)$ is a binomial-based estimator of the theoretical proportion

$$\Pi = \frac{\mu_1}{\mu_1+\mu_2} = \frac{\lambda_1 \times PT_1}{\lambda_1 \times PT_1 + \lambda_2 \times PT_2} = \frac{\theta \times PT_1}{\theta \times PT_1 + PT_2}.$$

In addition, the corresponding binomial-based CI, $(\Pi_{LOWER},\ \Pi_{UPPER})$, for the corresponding theoretical proportion is a CI for the theoretical quantity $(\theta \times PT_1)/(\theta \times PT_1 + PT_2)$. Since the CI is calculable from y_1 and y, and since PT_1 and PT_2 are known, we can back-calculate from

$$\Pi = \frac{\theta \times PT_1}{\theta \times PT_1 + PT_2},$$

to obtain

$$\{\theta_{LOWER},\ \theta_{UPPER}\} = \left\{ \frac{\Pi_{LOWER}}{1-\Pi_{LOWER}},\ \frac{\Pi_{UPPER}}{1-\Pi_{UPPER}} \right\} \div \frac{PT_1}{PT_2}.$$

Example 5.4 from A & B.

"Equal volumes $[V_1 = V_2 = V]$ of two bacterial cultures are spread on nutrient media and after incubation the numbers of colonies growing on the two plates are 13 and 31. We require confidence limits for the ratio of concentrations of the two cultures.

The estimated ratio is $(13/V) \div (31/V) = 0.4194$. From the Geigy tables a binomial sample with 13 successes out of 44 provides the following 95% confidence limits for Π: 0.1676 and 0.4520. Calculating $\Pi \div (1-\Pi)$ for each of these limits gives the following 95% confidence limits for the concentration ratio: $0.1676/0.8324 = \textit{0.2013}$ and $0.4520/0.5480 = \textit{0.8248}$.

The normal approximations described in section 4.4 can, of course, be used to obtain the CI for Π when the frequencies are not too small."

Example Breast cancer cases and person years of observation for women with tuberculosis repeatedly exposed to multiple x-ray fluoroscopies, and women with tuberculosis not so exposed [Boice and Monson, 1977][2]

	Radiation exposure		
	Yes	No	Total
Breast cancers	41	15	56
Person-years	28,010	19,017	47,027

Point estimate of Rate Ratio: $\hat{\theta} = (41/28,010) \div (15/19,017) = 1.86$

The observed proportion of exposed cases is $41/56 = 0.732$, with accompanying 95% binomial CI (0.596, 0.842). Thus,

$$\{RR_{LOWER}\ ,\ RR_{UPPER}\} = \left\{ \frac{0.596}{0.404}\ ,\ \frac{0.842}{0.158} \right\} \div \frac{28,010}{19,017} = \{1.00, 3.61\}.$$

[2]Example 11-1 from Rothman 1986, pp 156, Ch 11.

Example: Risk[3] of Motor Vehicle Crashes after Extended Shifts (24 hr) .[4]

	Extended Shifts	Nonextended Shifts
No. reported crashes	58	73
No. of commutes	54,121	180,289
Rate (per 1000 commutes)	1.07	0.40
RR: Rate ratio (95% CI•)	2.65 (1.87 to 3.74)	1.0

• 95% CI for RR (unmatched analysis):
$2.65 \div / \times \exp[1.96 \times (1/58 + 1/73)^{1/2}] = 2.7 \div / \times 1.41 = 1.87$ to 3.74.

Other methods:

• $X^2 = (58 - 30.25)^2/30.25 + (73 - 100.75)^2/100.75 = 33.11$, so $X = 5.75$.
95% **test-based** CI: $2.65^{1 \mp 1.96/5.75} = 1.90$ to 3.69.

• Based on conditional approach: from 58/131, CI for Π: 0.3488 to 0.5245, so

$$RR_{LOWER},\ RR_{UPPER}\} = \left\{ \frac{0.3488}{0.6512},\ \frac{0.5245}{0.4755} \right\} \div \frac{54,121}{180,289} = 1.78 \text{ to } 3.67.$$

Example: Efficacy Analyses of a Human Papillomavirus Type 16 L1 Virus-like-Particle Vaccine.[5]

• In the '*Primary per-protocol efficacy analysis*,' there were 0 persistent infections in 1084.0 w-y of vaccinated follow-up, versus 41 in 1076.9 w-y of placebo follow-up.

95% CI for proportion based on 0/41: 0 to 0.086.

$$\{RR_{LOWER},\ RR_{UPPER}\} = \left\{ \frac{0.0000}{1.0000},\ \frac{0.086}{0.914} \right\} \div \frac{1084.0}{1076.9} = 0 \text{ to } 0.093.$$

$$\{Efficacy_{upper},\ Efficacy_{lower}\} = 100\% \text{ to } 90.7\%$$

Article gave point est. of 100 percent, 95% CI of 90 percent to 100 percent.

[3]Point- and interval-estimates in article are based on within-person comparisons.
[4]table 1. N Engl J Med 352;2 www.nejm.org january 13, 2005
[5]Koutsky LA, N Engl J Med 2002;347:1645-51.

• In the '*secondary efficacy analysis*,' there were 6 and 68 'transient or persistent' infections.

95% CI for proportion based on 6/74: 0.0304 to 0.1681.

$$\{RR_{LOWER},\ RR_{UPPER}\} = \left\{\frac{0.0304}{0.9696},\ \frac{0.1681}{0.8319}\right\} \div \frac{1084.0}{1076.9} = 0.0311 \text{ to } 0.2007.$$

$$\{Efficacy_{upper},\ Efficacy_{lower}\} = 96.89\% \text{ to } 79.93\%$$

Article reported 97% to 80%.

The Statistical Analysis in the article stated (*italics by JH*):

> For all efficacy analyses, a point estimate of vaccine efficacy and the 95 percent confidence interval were calculated on the basis of the observed case split between vaccine and placebo recipients and the accrued person-time. The statistical criterion for success required that the lower bound of the two-sided 95 percent confidence interval for vaccine efficacy exceed 0 percent. For the primary analysis, this corresponds to a test (two-sided $\alpha = 0.05$) of the null hypothesis that the vaccine efficacy equals 0 percent. *An exact conditional procedure, which assumes that the numbers of cases in the vaccine and placebo groups are independent Poisson random variables*, was used to evaluate vaccine efficacy.

4 Sample Size Requirements for Comparison of Rates

4.1 Expected numbers of events required in Group 1 to give specified power, $\alpha = 0.05$

Relative Rate*	Expected events in Group 1 to yield power of ... 80%	90%	95%
0.1	10.6	14.3	17.6
0.2	14.7	19.7	24.3
0.3	20.8	27.9	34.4
0.4	30.5	40.8	50.4
0.5	47.0	63.0	77.8
0.6	78.4	105.0	129.6
0.7	148.1	198.3	244.8
0.8	352.8	472.4	583.2
0.9	1489.6	1994.5	2462.4
1.1	1646.4	2204.5	2721.6
1.2	431.2	577.4	712.8
1.4	117.6	157.5	194.4
1.6	56.6	75.8	93.6
1.8	34.3	45.9	56.7
2.0	23.5	31.5	38.9
2.5	12.2	16.3	20.2
3.0	7.8	10.5	13.0
5.0	2.9	3.9	4.9

* Ratio of incidence rate in Group 2 to incidence rate in Group 1.

Using a two-sided significance test with $\alpha = 0.05$.

The two groups are assumed to be of equal size (Breslow & Day more general)

Numbers taken from Table 3.2 in Chapter 3 "Study Size" in "Methods for Field Trials of Interventions against Tropical Diseases: A Toolbox" Edited by P.G. Smith and Richard H. Morrow. Oxford University Press Oxford 1991. (on behalf of the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases). *See Resources.*

Note that roles of Group 1 and 2 above are reversed from in Smith & Morrow text; See also Breslow NE and Day NE Vol II, Section 7.3.

4.2 Formulae for calculating study size requirements for comparison of rates using two groups of equal size:

From Table 3.4 of Morrow and Smith, with role of groups 1 & 2 reversed.

• Choosing study size to achieve adequate precision (§3.2 in text)

$$e_1 = (1.96/\log f)^2 \times \{(RR+1)/RR\}$$

e_1 = Expected no. of events in group 1
RR = (Rate in group 2) ÷ (Rate in group 1)

Gives 95% CI from $RR \div f$ to $RR \times f$

• Choosing study size to achieve adequate power (§4.2 in text)

$$\text{P-T} = (z_{\alpha/2} + z_\beta)^2 \times (R_2 + R_1) \;/\; (R_2 - R_1)^2$$

P-T = Person-time in each group
R_i = Rate in group i

$z_{\alpha/2} = 1.96$ for $\alpha = 0.05$ (2-sided)

Power:	80%	90%	95%
z_β :	0.84	1.28	1.64

5 Extra-Poisson Variation

5.1 e.g.: Daylight Savings Time & Traffic Accidents

To the Editor[6]: It has become increasingly clear that insufficient sleep and disrupted circadian rhythms are a major public health problem. For instance, in 1988 the cost of sleep-related accidents exceeded $56 billion and included 24,318 deaths and 2,474,430 disabling injuries.[1] Major disasters, including the nuclear accident at Chernobyl, the Exxon Valdez oil spill, and the destruction of the space shuttle Challenger, have been linked to insufficient sleep, disrupted circadian rhythms, or both on the part of involved supervisors and staff.[2,3] It has been suggested that as a society we are chronically sleepdeprived[4] and that small additional losses of sleep may have consequences for public and individual safety.[2]

We can use noninvasive techniques to examine the effects of minor disruptions of circadian rhythms on normal activities if we take advantage of annual shifts in time keeping. More than 25 countries shift to daylight savings time each spring and return to standard time in the fall. The spring shift results in the loss of one hour of sleep time (the equivalent in terms of jet lag of traveling one time zone to the east), whereas the fall shift permits an additional hour of sleep (the equivalent of traveling one time zone to the west). Although one hour's change may seem like a minor disruption in the cycle of sleep and wakefulness, measurable changes in sleep pattern persist for up to five days after each time shift.[5] This leads to the prediction that the spring shift, involving a loss of an hour's sleep, might lead to an increased number of "microsleeps," or lapses of attention, during daily activities and thus might cause an increase in the probability of accidents, especially in traffic. The additional hour of sleep gained in the fall might then lead conversely to a reduction in accident rates.

We used data from a tabulation of all traffic accidents in Canada as they were reported to the Canadian Ministry of Transport for the years **1991 and 1992** by all 10 provinces. A total of 1,398,784 accidents were coded according to the date of occurrence. **Data for analysis were restricted to the Monday preceding the week of the change due to daylight savings time, the Monday immediately after, and the Monday one week after the change, for both spring and fall time shifts.** Data from the province of Saskatchewan were excluded because it does not observe daylight savings time. The analysis of the spring shift included 9593 accidents and that of the fall shift 12,010. The resulting data are shown in Figure 1.

[6] Correspondence: New Engl. J of Medicine: Vol. 334:924-925 April 4, 1996.

The loss of one hour's sleep associated with the spring shift to daylight savings time increased the risk of accidents. The Monday immediately after the shift showed a relative risk of 1.086 (95 percent confidence interval, 1.029 to 1.145; $\chi^2 = 9.01$, 1 df; P<0.01). As compared with the accident rate a week later, the relative risk for the Monday immediately after the shift was 1.070 (95 percent confidence interval, 1.015 to 1.129; $\chi^2 = 6.19$, 1 df; P<0.05). Conversely, there was a reduction in the risk of traffic accidents after the fall shift from daylight savings time when an hour of sleep was gained. In the fall, the relative risk on the Monday of the change was 0.937 (95 percent confidence interval, 0.897 to 0.980; $\chi^2 = 8.07$, 1 df; P<0.01) when compared with the preceding Monday and 0.896 (95 percent confidence interval, 0.858 to 0.937; $\chi^2 = 23.69$; P<0.001) when compared with the Monday one week later. Thus, **the spring shift to daylight savings time, and the concomitant loss of one hour of sleep, resulted in an average increase in traffic accidents of approximately 8 percent**, whereas the fall shift resulted in a decrease in accidents of approximately the same magnitude immediately after the time shift.

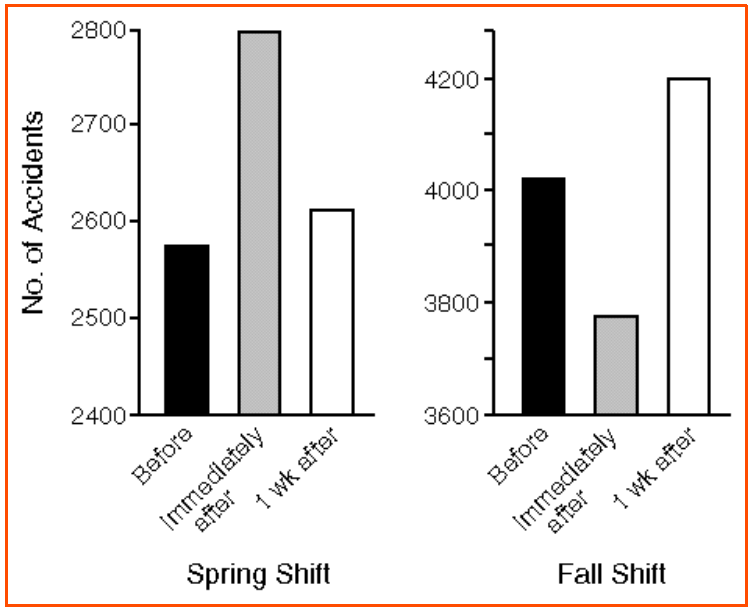

Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylight Savings Time for the Years 1991 and 1992. *There is an increase in accidents after the spring shift (when an hour of sleep is lost) and a decrease in the fall (when an hour of sleep is gained).*

These data show that small changes in the amount of sleep that people get can have major consequences in everyday activities. The loss of merely one hour of sleep can increase the risk of traffic accidents. It is likely that the effects are due to sleep loss rather than a nonspecific disruption in circadian rhythm, since gaining an additional hour of sleep at the fall time shift seems to decrease the risk of accidents.

Stanley Coren, Ph.D.
University of British Columbia
Vancouver, BC V6T 1Z4, Canada

References:

1. Leger D. The cost of sleep-related accidents: a report for the National Commission on Sleep Disorders Research. Sleep 1994;17:84-93. [Medline] 2. Coren S. Sleep thieves. New York: Free Press, 1996. 3. Mitler MM, Carskadon MA, Czeisler CA, Dement WC, Dinges DF, Graeber RC. Catastrophes, sleep, and public policy: consensus report. Sleep 1988;11:100-109. 4. Webb WB, Agnew HW Jr. Are we chronically sleep deprived? Bull Psychonom Soc 1975;6:47-8. 5. Monk TH, Folkard S. Adjusting to the changes to and from daylight saving time. Science 1976;261:688-9.

* * * * * * *

NEJM Volume 339:1167-1168 October 15, 1998

Effects of Daylight Savings Time on Collision Rates

To the Editor: The results of a recent Canadian study call into question Coren's findings that motor vehicle crashes increase by 8 percent following the change to daylight savings time and decrease by 7 percent after the change to standard time.[1] The study extended Coren's analysis, using the same data source. First, data from the days between the Monday preceding the time change and the Monday one week afterward were analyzed. Second, Coren's hypothesis was statistically tested with data from the years 1984 to 1993, to evaluate the significance of any differences obtained.

A graphical analysis (Figure 1) indicated that there were no peaks on the Mondays after the **(Spring) change to daylight savings time** or troughs on the Mondays after the return to standard time, and no corresponding return to base-line values after each transition. *The results of a paired t-test with pooled national data failed to reach significance (P=0.5), with a mean rate of 129.8 for the Monday one week before and for the Monday immediately after the change to daylight savings time (95 percent confidence interval for the difference between the means, -12.12 to +12.43).* An analogous paired

t-test with pooled national data found no difference, with the mean rate one week following the change equal to 130.1 (95 percent confidence interval for the difference between the means, -13.12 to +13.84).

Next, the mean motor vehicle crash rates for the Monday one week before the **(Fall) change to standard time** were compared with those for the Monday immediately after the change and *showed a significant increase (160.8 vs. 188.5; 95 percent confidence interval for the difference between the means, 6.6 to 48.4; t=2.66; P<0.01).* This result is inconsistent with Coren's hypothesis. A paired t-test showed that the mean rate of 188.5 for the Monday immediately after the change was not significantly different from the mean rate of 186.5 for the Monday one week after the change (95 percent confidence interval for the difference between the means, 14.6 to +12.7; t=0.16; P=0.4).

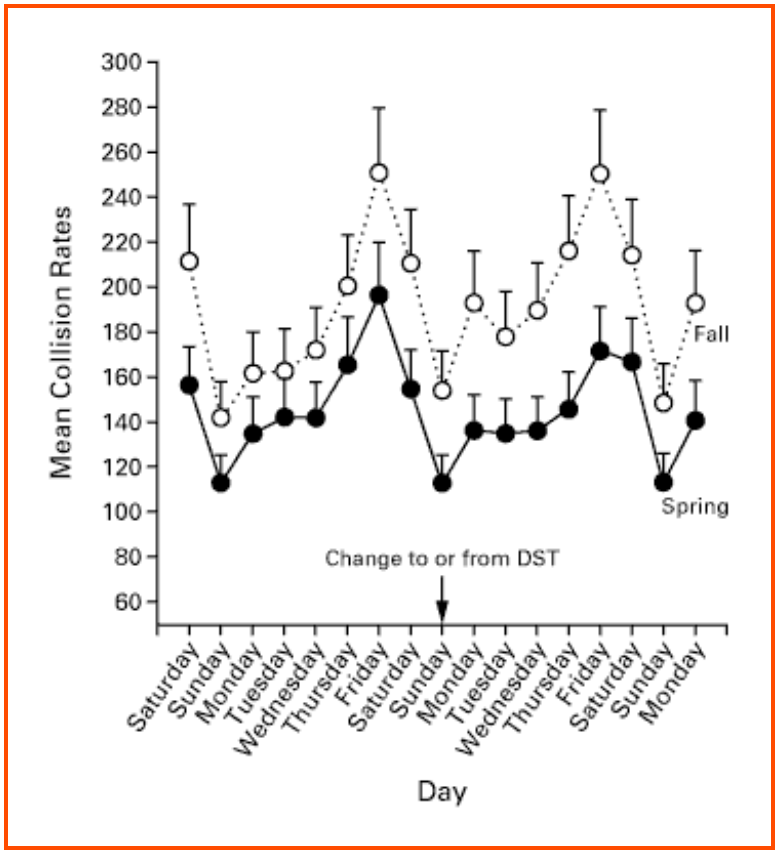

Figure 1. Effect of the Change to (**solid circles**) or from (open circles) Daylight Savings Time (DST) on the Mean (+SE) Collision Rates.

Thus, the results of both a graphical analysis and the **variability estimation of 10 years of data** failed, as had an earlier study,[2] to support Coren's hypothesis. The effects reported by Coren may stem from the increased number of vehicles on the road and the increased number of kilometers traveled in the extra daylight hour in the spring, rather than from the minor disruption in circadian rhythm induced by the loss of one hour of sleep.

Alex Vincent, Ph.D.
Transport Canada
Montreal, QC H3B 1X9, Canada

References

1. Coren S. Daylight savings time and traffic accidents. N Engl J Med 1996;334:924- 924. [Free Full Text] 2. Stewart DE. The implications of an early return to day-light saving time in Canada: impact on road safety. Technical memorandum. Road Safety and Motor Vehicle Regulation TMSE 8503. Ottawa, Ont.: Transport Canada, 1985.

Dr. Coren replies:

To the Editor: In my study of the effects of daylight savings time on traffic accidents, I found increased accident rates on the Monday after the spring shift in time and decreased rates in the fall. I interpreted this in terms of sleep time lost or gained. Vincent uses a larger data base than that available to me and fails to replicate these results. **Unfortunately, Vincent's analyses are based on t-tests of annual counts, rather than more sensitive[7] pooled relative-risk measures.** More important, analysis of recent data from larger data banks gives me reason still to believe that the shift to daylight savings time in the spring is associated with an increased risk of accidents, although the rebound reduction in accidents in the fall may be more problematic.

In an extension of the original study, I obtained data from the National Highway Traffic Safety Administration on all (366,910) deaths in the United States due to traffic accidents for the years 1986 through August 1995.[1] Data were cumulated over the 10-year period. Contrasting traffic fatalities for the Monday immediately after the spring shift to daylight savings time with the pooled frequency for the Mondays preceding and following that date shows the expected significant increase, with a relative risk of 1.17 (95 percent confidence interval, 1.07 to 1.29; $\chi^2 = 10.83$, 1 df; P<0.001). The magnitude of this shift

[7] Coren's method is **too sensitive**. His 'too small SE' is based on a 'homogeneous Poisson' model that does not fit the data.

is larger than in my first study, amounting to 17.2 percent. The fall time shift, however, was associated with an insignificant reduction in traffic deaths (2.6 percent), with a relative risk of 0.97 (95 percent confidence interval, 0.89 to 1.07; $\chi^2 = 0.29$, 1 df; P not significant).

This result is similar to that of studies based on accidental deaths not related to traffic accidents.[2,3] For example, I looked at every accidental death in the United States that was reported to the National Center for Health Statistics for the years 1986 through 1988.[4] Since over 80 percent of accident-related deaths occur within four days after the accident, data for the analysis were restricted to the first four workdays immediately following the change to daylight savings time and the first four workdays in the week preceding and the week after the change. There were 8429 accidental deaths in the spring-shift analysis and 8771 in the fall. The interval immediately following the spring shift showed a 6.6 percent increase in accidental deaths (relative risk, 1.07; 95 percent confidence interval, 1.01 to 1.11; $\chi^2 = 5.52$, 1 df; P<0.05). The fall shift, however, was associated with a nonsignificant 1.5 percent decrease (relative risk, 0.99; 95 percent confidence interval, 0.922 to 1.021; $\chi^2 = 1.34$, 1 df; P not significant). These data are consistent with the hypothesis that a small decrease in the duration of sleep can increase one's susceptibility to accidents. Although work schedules accentuate the loss of sleep after the spring shift to daylight savings time, *the absence of a reduction in accidents in the fall may reflect the fact that many people do not take advantage of the hour gained to extend their sleep.*

Stanley Coren, Ph.D.
University of British Columbia
Vancouver, BC V6T 1Z4, Canada

References

1. National Center for Statistics and Analysis. Fatal accident reporting system, 19861995. (Computer files.) Washington, D.C.: National Highway Traffic Safety Administration, 19861995. 2. Coren S. Sleep thieves. New York: Free Press, 1996. 3. Coren S. Accidental death and the shift to daylight savings time. Percept Motor Skills 1996;83:921-2. 4. National Center for Health Statistics. Multiple cause of death, 1986, 1987, 1988. (Computer files.) Hyattsville, Md.: Department of Health and Human Services, 19861988.

5.2 Extra-Poisson variation: Notes by JH

The variability uncovered by Vincent is an example of "extra-Poisson" (i.e., "larger than Poisson") variation.

The derivation of the Poisson distribution assumes that there is the *same* infinitely small probability of an event for each person-moment, **or** that we are dealing with the sum of a very large – but *fixed* number (distribution) – of Bernoulli r.v.'s, each with *its own* very small (but unknown to us) probability of being positive, i.e., both $\sum_{i=1}^{i=10^5} Bernoulli(\mu/10^5)$ and $\{\sum_{i=1}^{i=5\times10^4} Bernoulli(2\mu/10^5) + \sum_{i=1}^{i=5\times10^4} Bernoulli(0)\}$ will have a Poisson(μ) distribution, provided $\mu << 10^5$. As an example, we could use the same Poisson($\mu = 151.5$) distribution to describe *year to year variation* in the numbers of new breast cancers, whether they arise from a combined sample of

- 0.5×10^5 65 year old men, in whom the average incidence is $3/10^5 py$, and 0.5×10^5 65 year old women in whom it is $300/10^5 py$, [8]
- 0.9925×10^5 45 year old women, in whom the average incidence is $150^5 py$, and 0.0075×10^5 75 year old women in whom it is $350/10^5 py$,
- 1×10^5 46 year old women, in whom the average incidence is $151.5^5 py$,
- 0.5×10^5 46 year old lower-risk women, in whom the average incidence is $75.75^5 py$, and 0.5×10^5 46 year old higher-risk women in whom it is $303/10^5 py$.

The yearly numbers emanating from a Poisson($\mu = 151.5$) series would stay mainly within the range 125 to 175. The **key** is the 'fi**xed portfolio**.'

The above input rates are based on rates seen recently in the UK. But now, imagine that the observed number of cases for one year was derived from the U.K, for the next year were from the U.S.A, the next from Japan, the next from Equador, etc... but that you did not know that. For example, suppose that a not very careful data-processor reported on a different random source each year without telling you, or mixed up breast cancer and colon cancer, or used data on new cases of influenza instead, or used numbers of persons dying in small plane crashes, or reported on a case series where there was large turnover in, and little supervision of, the persons who coded the diagnosis.

All of these would lead to considerably more year to year (or unit to unit) variation than would be predicted by the Poisson model.

[8] from `http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/`

We refer to this extra variation as "extra-Poisson" variation. It is usually caused by **external** factors, often unknown. Sometimes, we might be able to understand why it occurred, and possibly collect some (but probably not all) the variables responsible. For example, the yearly fluctuations in the recorded numbers of traffic accidents could be caused by such things as yearly (or sudden) variations in (a) the numbers of kilometers driven, which might be a function of the cost of gasoline, or the strength of the economy, or the price of a Metro pass (b) the degree of police surveillance and enforcement (c) weather (d) coding and data-processing practices, etc etc.. Similarly one would see considerably greater than Poisson variation in the numbers of cases of diseases that (i) are contagious (ii) can be screened for (ii) receive increased attention or publicity (eg new cases of attention deficit disorder, autism, erectile dysfunction...). And of course, the numbers who are truly at risk can change too – for many behaviors, such as motor-cycle – or helmet, or protective ski equipment – use, it is not easy to document changes in the yearly denominators.

JH suspects that weather was one of the major reasons for the considerably greater than Poisson yearly variations uncovered by Vincent is *the weather*. This *affects both the denominator* (the number of cars on the road) and the *crash probability per car*. It turns out that in Coren's 2 years, the Monday before the "Spring ahead" weekend in 1991 was Monday April 1, which was also Easter Monday, a holiday for many people. It was not that there was a greater number of accidents on Monday April 8th (just after the switch to DST; rather it was that there were fewer on Easter Monday. Whereas Coren was probably aware of this, to the regular reader, this was an unknown source of variation.

Clinical trial context: several events per person

Clinical trials of interventions for chronic diseases such as asthma, COPD, and MS often use the number (y) of attacks – or exacerbations or hospitalizations – a patient experiences over a fixed amount of follow-up time (PT) as the 'response' variable.

Unfortunately, some data-analysts then pool all of the follow-up time in the one arm (with say n patients) to get $PT = \sum PT_i$, pool all of the events in that arm to get $c = \sum y_i$, and treat this total number (c) of events or 'cases' as a Poisson random variable.

If we were dealing with very few events per person (if most y's were 0, a few were 1, even fewer were 2, etc , as with the needlestick injuries or car crashes examples), the implications of such an approach would not be serious. But when there is considerable variability in the event rates across persons, and many of the y's are in the double digits, c has far more variability (from one possible sample of n to another) that would be predicted by treating it as Poisson variable.

To see this, think of the count y_i for the ith randomly selected person in the study as arising from a 2-stage process. This person was chosen for the study from a pool of persons. Each person in the pool has his/her own expected rate λ, and thus his/her own expected number of events over the follow-up period in question (for simplicity, we pretend that all study subjects will be followed for the same amount of time: we take it be 1 PT unit). Thus, the observed number y_i of events experienced by person i is governed by his/her λ_i. Say that the distribution of λ's in the pool has an overall mean μ_λ and variance σ^2_λ

What then is the expected value, and variance, of the random variable y_i?

We need to take these over the two stages, the selection of the λ_i and the possible realizations given λ_i.

The expectation is $E[y_i] = E[\ E[y_i \mid \lambda_i]\ ] = E[\ \lambda_i\ ] = \mu_\lambda$

The variance is in two parts

$$Var[y_i] = E[\ Var[y_i \mid \lambda_i]\ ] + Var[\ E[y_i \mid \lambda_i]\ ] = E[\ \lambda_i\ ] + Var[\ \lambda_i\ ] = \mu_\lambda + \sigma^2_\lambda$$

Thus, the 'unit variance' is larger than the Poisson variance whenever there is heterogeneity in the λ's, i.e. when $\sigma^2_\lambda > 0$.

This two stage (or hierarchical) model is often written as

$$\lambda_i \sim ?(\ ),$$
$$y_i \mid \lambda_i \sim Poisson(\lambda_i).$$

How is it that the variation from say year to year in the breast cancer example might be governed by the Poisson law, but the variation in the c from one possible sample of n to other sample of n is not?

The *answer* lies in **whether the 'portfolio' is fixed or random**.

In the breast cancer example, it was fixed; in the selection of a sample for an rct, it is random. Even without any effect of the intervention, the possible variation of the c in each treatment arm from $n \times \mu_\lambda$ could be much greater than that suggested by assuming $c \sim Poisson(\sum \mu_\lambda = n \times \mu_\lambda)$.

Same concept: "extra-Binomial" variation

The same issues apply to "extra-Binomial" variations. For example, while one would expect the numbers of left-handers in a class of 30 to vary according to a Binomial with a reasonably steady π, we would not expect the same regularity in the numbers wearing jeans, or beards, or dyed hair, or sporting iPods. For some of these behaviours, it might be possible to document smooth trends over time, or as a function of the age/sex distribution, but for others we might have very few explanations for the extra-binomial variation.

A classic example of "extra-Binomial" variation is in the context of cluster sampling or cluster-randomized trials: for example, in the rubella-protection prevalence study, the proportions of persons who dropped out of exercise classes, etc. Cochran has a nice example in his sampling textbook: the proportion of persons who have visited a physician visits in the last year, estimated from a household survey.

Not all non-Poisson or non-Binomial variation is 'extra.' [most is!]

There are a small number of situations where there is *less tha*n the model-based variation. For example, if McDonald's puts an average of $\mu = 16$ olives per pizza, one would expect McDonald's uniformity to result in a SD, from pizza to pizza, of far less than the $\sigma = \mu^{1/2} = 4$ predicted by the Poisson model. Likewise, whereas the proportion (π) of males in the population can be estimated from a household survey, the distribution of the number (y) of males in households of say $n = 4$ persons is considerably less than the $\sigma_y = \{4 \times \pi \times (1 - \pi)\}^{1/2}$ predicted by the Binomial.

What to do about extra-Poisson and extra-Binomial variation?

- It is likely to exist. Check for it.
- Model it (and thus remove some of it) using the explanatory variables at hand.
- If considerable extra variation persists, despite the attempts to understand it via the available covariates, do not use the (too narrow) model-based SEs; instead use empirical SEs that reflect the observed, not the model-predicted, unit variation. In effect, that's what Vincent did for the with-year differences in accident rates. Or use parametric models that allow extra variation. For Poisson-like counts, but with extra-Poisson variation, one 'next model up' is the negative binomial probability distribution: it has two parameters rather than the one that governs the binomial. Another (should be equivalent) to model the counts as a mixture of Poisson random variables, where the μ's have a gamma distribution. In principle, more complex mixing distributions could be used to handle the additional variance.

POISSON, Siméon Denis 1781-1840

`http://www.york.ac.uk/depts/maths/histstat/people/sources.htm#p`

0 Exercises

0.1 ms - Distribution of $y_1 \mid y_1+y_2$, when $y_1 \sim Poisson(\mu_1)$ and (independently) $y_2 \sim Poisson(\mu_2)$

Exercise 4.15, Casella & Berger, p194.

0.2 ms - Two ways of deriving the Negative Binomial Distribution

• As a sum of k i.i.d. Geometric r.v.'s

The '**zero**-origin' version of the Geometric (G) probability distribution is the probability distribution of the number Y of failures **before** the first success, supported on the set $\{0, 1, 2, 3, ...\}$, when the probability of success on each trial is π.

$$Prob_G[Y = y] = (1 - \pi)^y \times \pi.$$

With this zero-origin version, the Negative Binomial (NB) probability distribution with parameters k and π is the probability distribution of the number Y of failures **before** the k-th success, again supported on the set $\{0, 1, 2, 3, ...\}$.

Exercise: Show that, so defined,

$$Prob_{NB}[Y = y] \;=\; {}^{(y+k-1)}C_{k-1} \times (1 - \pi)^y \times \pi^k.$$

and (from the pmf, or directly from the definition as a sum) find its expectation and variance.

• The negative binomial distribution can also be thought of as a continuous mixture of Poisson (P) distributions where the mixing distribution of the Poisson rate [or mean–JH] is a gamma distribution. Formally, this means that the mass function of the negative binomial distribution can also be written as

$$\begin{aligned} Prob_{NB}[Y = y] &= \int_0^\infty pmf_P[y \mid \mu] \times pdf_\Gamma[\mu \mid k, (1 - \pi)/\pi \,] d\mu \\ &= \ldots \\ &= \frac{\Gamma[k + y]}{y!\ \Gamma[k]} \times \pi^k \times (1 - \pi)^y. \end{aligned}$$

Because of this, the negative binomial distribution is also known as the gamma-Poisson (mixture) distribution.

Exercise: fill in the omitted steps.

(above text from `http://en.wikipedia.org/wiki/Negative_binomial_distribution`)

From `R` documentation... [see `nbinom` in `R`]

A *negative binomia*l distribution can arise as a mixture of Poisson distributions with mean distributed as a Γ (gamma) distribution with scale parameter $(1-\pi)/\pi$ and shape parameter k. (This definition allows non-integer values of k.) In this model $\pi = scale/(1+scale)$, and the mean is $k \times (1-\pi)/\pi$.

The alternative parametrization (often used in ecology) is by the mean μ_{NB}, and k, the dispersion parameter, where $\pi = k/(k+\mu)$. The variance is $\mu_{NB} + \mu_{NB}^2/k$ in this parametrization or $k \times (1-\pi)/\pi^2$ in the first one.

Overdispersed Poisson: The negative binomial distribution, especially in its alternative parametrization described above, can be used as an alternative to the Poisson distribution. It is especially useful for discrete data over an unbounded positive range whose sample variance exceeds the sample mean. If a Poisson distribution is used to model such data, the model mean and variance are equal. In that case, the observations are overdispersed with respect to the Poisson model. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean...

0.3 ms - Sample size formulae for comparison of rates

See section 3.1 of the Notes. Derive the "Expected numbers of events required ... specified power" from 'scratch,' using normal approximations to the Poisson distributions of the observed numbers of events in the two (equal) amounts of population-time.

0.4 Extended Work Duration and the Risk of Self-reported Percutaneous Injuries in Interns

Refer to rows 2 and 3 of Table 3. in this article, by Ayas et al. in JAMA on Sept 6 of 2006. [Resources - Intensity]

1. Manually calculate ORs and 95% CIs, and repeat by computer software.

2. 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.

3. 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 a future homework).

4. Is OR the correct term for the ratio being estimated here?

0.5 John Snow's "Grand Experiment"

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

So, the *denominators* and *numerators* were...

	Water Source	
	Dirtier	Cleaner
No. of houses with	40,046	26,107
No. of fatal attacks	286	24

1. Calculate a 95% CI to accompany the rate ratio.

2. 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' (also known as 'shoe-leather epidemiology' to determine the water source of each of 40046 + 26107 = 66153 houses!

   Use `R (or SAS or SPSS)` to form a sample of 100 houses that could form a '*denominator series.*' and thus to provide a point and interval estimate of the relative sizes of the two sources of water. Use the *case series* assembled by Snow, and the (armchair / virtual / desktop) *denominator series* obtained by you and `R`.

3. Repeat this virtual epidemiology, but using denominator series of 300, 600, and 2000. Comment on the estimates.

4. Explain to a journalist why are the CI's based on the virtual denominator series are wider than the one based on the actual "return which was made to Parliament"?

0.6 A population-based study of measles, mumps, and rubella vaccination and autism

Background[9]: It has been suggested that vaccination against measles, mumps, and rubella (MMR) is a cause of autism.

Methods: We conducted a retrospective cohort study of all children born in Denmark from January 1991 through December 1998. The cohort was selected on the basis of data from the Danish Civil Registration System, which assigns a unique identification number to every live-born infant and new resident in Denmark. MMR-vaccination status was obtained from the Danish National Board of Health. Information on the childrens autism status was obtained from the Danish Psychiatric Central Register, which contains information on all diagnoses received by patients in psychiatric hospitals and outpatient clinics in Denmark. We obtained information on potential confounders from the Danish Medical Birth Registry, the National Hospital Registry, and Statistics Denmark.

Results: Of the 537,303 children in the cohort (representing 2,129,864 person-years), 440,655 (82.0 percent) had received the MMR vaccine. We identified 316 children with a diagnosis of autistic disorder and 422 with a diagnosis of other autistic-spectrum disorders. After adjustment for potential confounders, the relative risk of autistic disorder in the group of vaccinated children, as compared with the unvaccinated group, was 0.92 (95 percent confidence interval, 0.68 to 1.24), and the relative risk of another autistic-spectrum disorder was 0.83 (95 percent confidence interval, 0.65 to 1.07). There was no association between the age at the time of vaccination, the time since vaccination, or the date of vaccination and the development of autistic disorder.

Conclusions: This study provides strong evidence against the hypothesis that MMR vaccination causes autism.

• No. of Cases of autism (numerators) among children who did / did not receive MMR vaccination ...

Vaccinated		
Yes (1)	No (0)	
263	53	316 (**Cases**)

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

Vaccinated		
Yes (1)	No (0)	
1.65m cy	0.48m cy	2.13m cy (**Denominators**)

• Crude* Rate Ratio ...

Vaccinated				
Yes (1)	No (0)	**Rate Ratio (RR)**	ME	**95% CI for RR**
$\frac{263}{1.65m\ cy}$	$\frac{53}{0.48m\ cy}$	1.44*	1.34	1.07 to 1.93

$ME = \exp[1.96 \times \{1/263 + \ 1/53\}^{1/2}]$;
$RR_{lower} = 1.44 \div 1.34$; $RR_{upper} = 1.44 \times 1.34$.

* The reason for the large difference between the crude (1.44) and adjusted (0.92, 95% CI 0.68 to 1.24) rate ratios will be discussed when we come to confounding. The crude ratio in this example is simply for didactic purposes.

1. Hand-calculate the CI's for the 1.44 ratio by your usual ± way, and compare the 'work' involved with that in the "multiplied-by/divided-by" method – ie be a 'hand-calculator consultant' to Rothman.
2. Which method do you prefer? (if you have software that does it for you, this is merely a heuristic issue!)

[9] Madsen KM et al. N Engl J Med 2002;347:1477-82.

0.7 Effect of Raloxifene on Risk of BrCa in Postmenopausal Women

Context: Raloxifene hydrochloride is a selective estrogen receptor modulator that has antiestrogenic effects on breast and endometrial tissue and estrogenic effects on bone, lipid metabolism, and blood clotting.

Objective: To determine whether women taking raloxifene have a lower risk of invasive breast cancer.

Design and Setting: The Multiple Outcomes of Raloxifene Evaluation (MORE), a multicenter, randomized, double-blind trial, in which women taking raloxifene or placebo were followed up for a median of 40 months (SD, 3 years), from 1994 through 1998, at 180 clinical centers composed of community settings and medical practices in 25 countries, mainly in the United States and Europe. Participants A total of 7500 postmenopausal women, younger than 81 (mean age, 66.5) years, with osteoporosis. Women who had a history of breast cancer or who were taking estrogen were excluded.

Intervention: Raloxifene, 60 mg, 2 tablets daily; or raloxifene, 60 mg, 1 tablet daily and 1 placebo tablet; or 2 placebo tablets.

Main Outcome: Measures New cases of breast cancer, confirmed by histopathology. Deep vein thrombosis or pulmonary embolism were determined by chart review.

Results: Thirteen cases of breast cancer were confirmed among the 5000 women assigned to raloxifene vs 26 among the 2500 women assigned to placebo (relative risk [RR], 0.25; 95% confidence interval [CI], 0.13-0.49; Chi- Square = 19.5, P¡.001). To prevent 1 case of breast cancer, 128 women would need to be treated. Raloxifene increased the risk of venous thromboembolic disease (RR, 3.0; 95% CI, 1.5-6.1), but did not increase the risk of endometrial cancer (RR, 0.8; 95% CI, 0.2-2.7).

Conclusion: Among postmenopausal women with osteoporosis, the risk of invasive breast cancer was decreased by 75% during 3 years of treatment with raloxifene.

Notes: (1) The numbers in the abstract have been "rounded" to make calculations easier. (2) The data from the 2 regimens were combined in the abstract. Thus, twice as many women received raloxifene as placebo.

Exercise:

1. Show how the authors calculated that "128 would need to be treated"
2. Reproduce the CI accompanying the RR of 0.25.
3. Would a test-based CI to give close to the same CI? Compute it and see.
4. If 3750 women each had been allocated to raloxifene & placebo, and the cancer rates been the same, would CI be narrower, wider or same?
5. Balance of benefits and risks: the abstract does not to report the numbers of thromboembolic disease, only the RR of 6 and the CI. From the information provided, determine – analytically or by trial and error – how many cases there were [Hint: $\log 1.5 = 0.4; \log 3 = 1.1; \log 6.1 = 1.8$; and use 2 as an approximation to 1.96]
6. In Table 2 of the paper, the authors report 15,000 and 7,500 women years of follow-up in the two groups. Does using these rather than the "person" denominators in the abstract change the CI's? Why/why not?