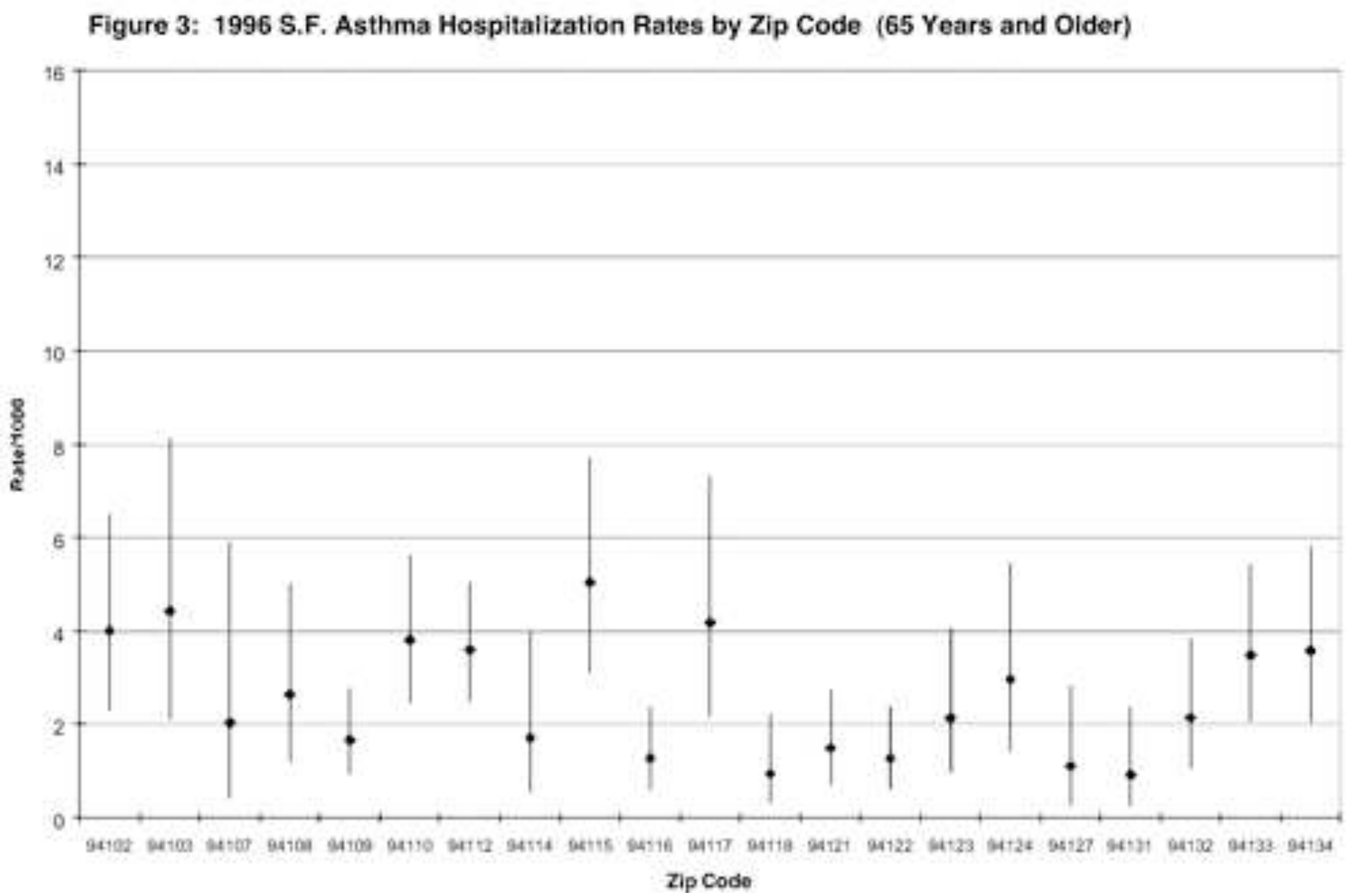
E H I B

Overlapping confidence intervals are not a statistical test

Daniel Smith
Environmental Health Investigations Branch
California Department of Health Services

26th Annual Institute on Research and Statistics
March, 2005

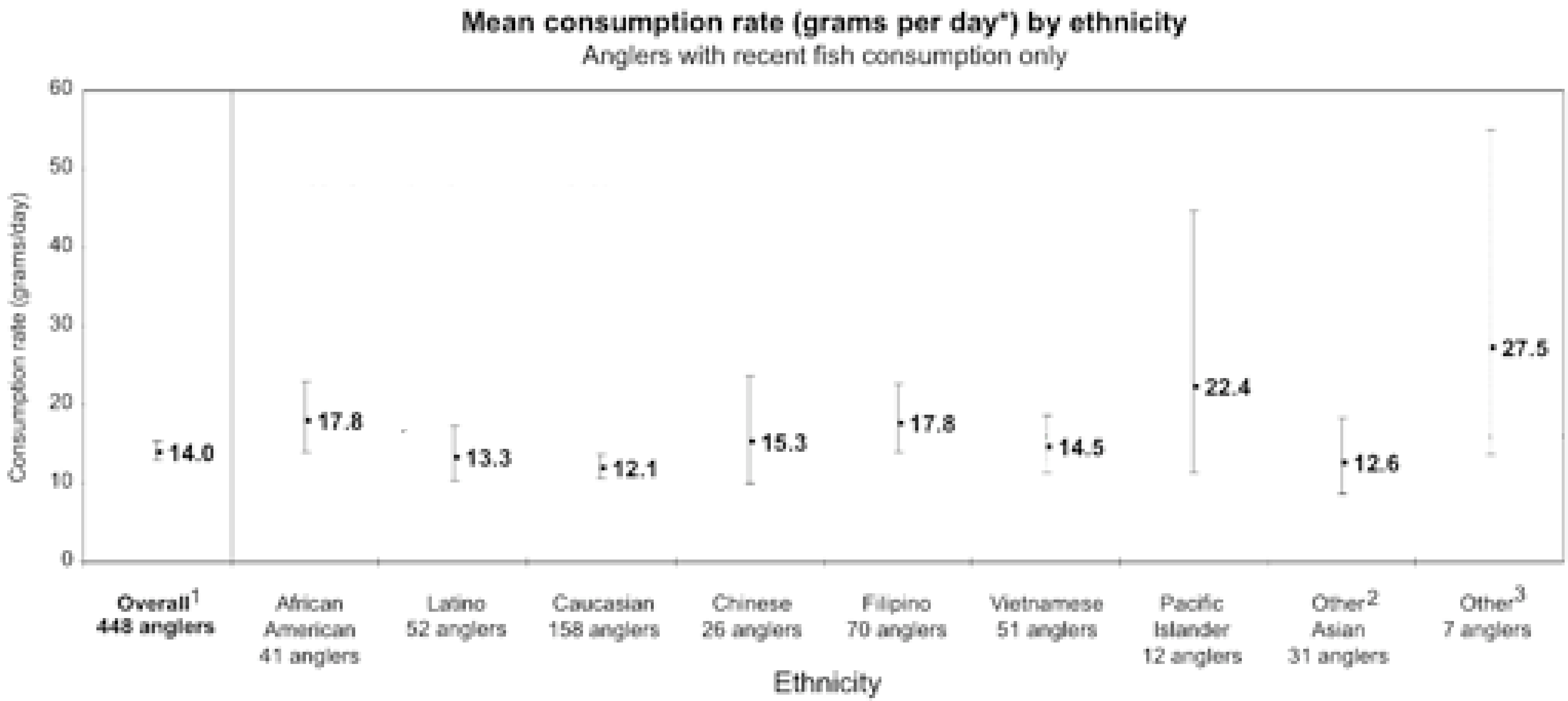
Some examples…



E H I B

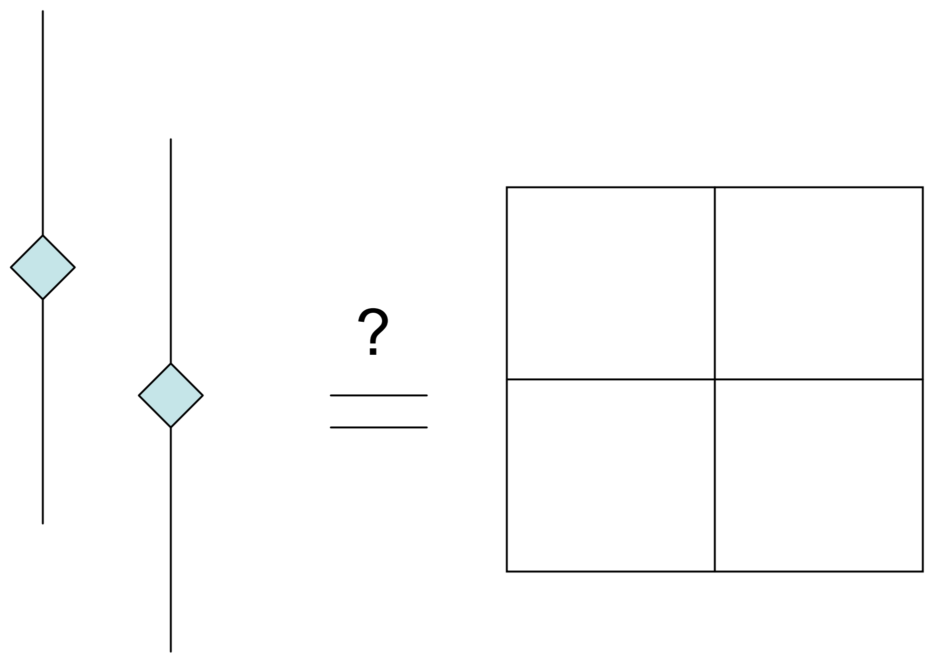
Daniel Smith
California Dept of Health Services

EHIB study of fish consumption…



Daniel Smith
California Dept of Health Services

Are the results of the overlap test consistent with the equivalent traditional test?



Daniel Smith
California Dept of Health Services

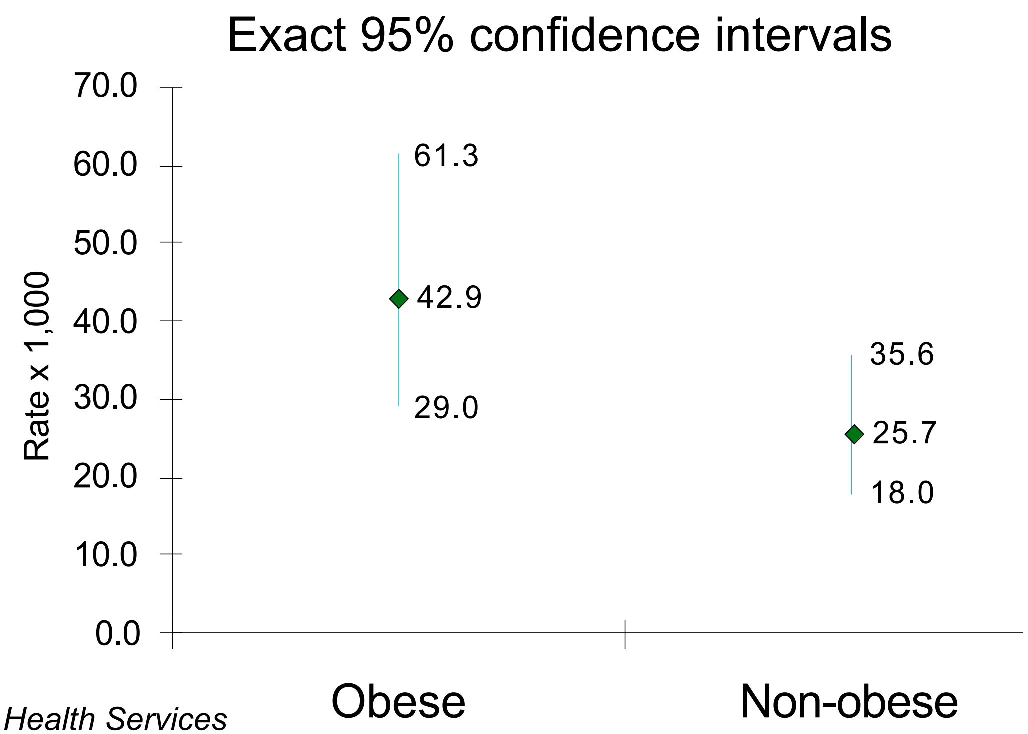
Example 1: Mortality rates among elderly

from Kleinbaum, Kupper, and Morgenstern, 1982, p.287.

	Obese	Non-obese
Deaths	30	36
Person-years	699	1399

$\chi^2 = 4.38$
P value = 0.036

RR=1.67
95% CL: 1.03-2.71



(CI Overlap)

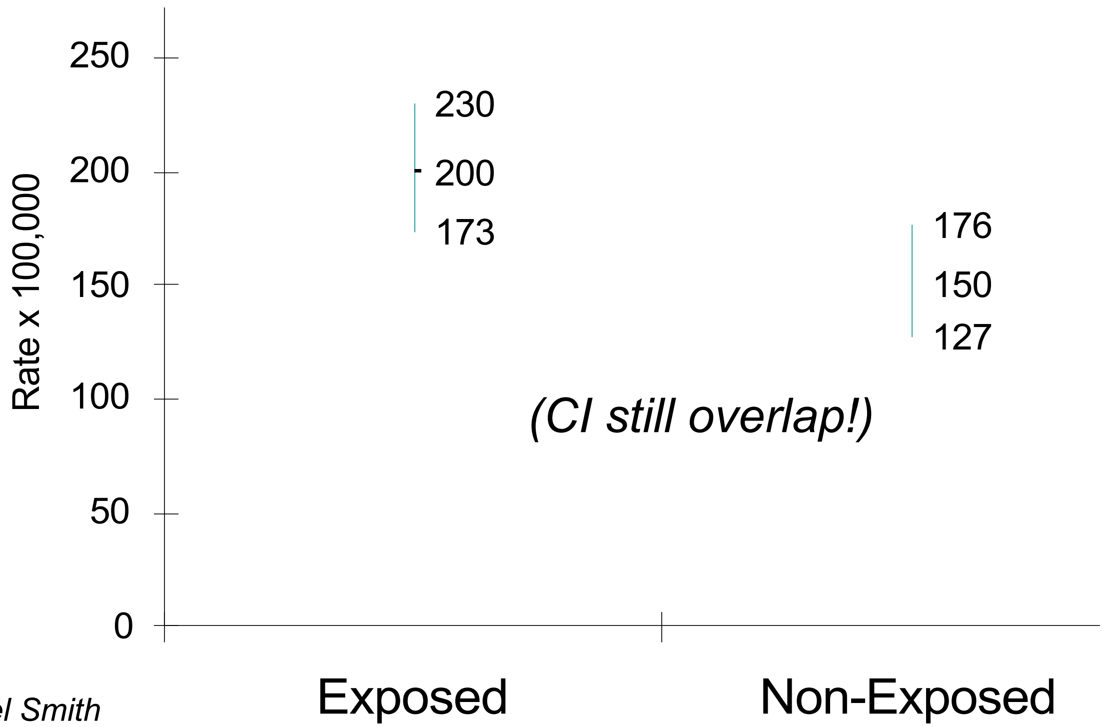
Daniel Smith
California Dept of Health Services

Example 2: Big Numbers

	Exposed	Non-Exposed
Cases	200	150
Person-years	100,000	100,000

$\chi^2 = 7.14$
P value = 0.0075

RR=1.33
95% CL: 1.08-1.65



Daniel Smith
California Dept of Health Services

The problem is…

- Confidence intervals are too wide
- Confidence intervals are calculated only using the data in each rate
- Confidence interval around *one* rate is not a comparative measure between *two* rates

ANOVA has the concept of Least Significant Difference (LSD), based on the confidence interval for the difference of two means

$$+Z * SE_{\overline{Y}_1-\overline{Y}_2}$$

$$\overline{Y}1 - \overline{Y}2$$

$$-Z * SE_{\overline{Y}_1-\overline{Y}_2}$$

$$\overline{Y}2$$

$$\frac{Z * SE_{\overline{Y}_1-\overline{Y}_2}}{2}$$

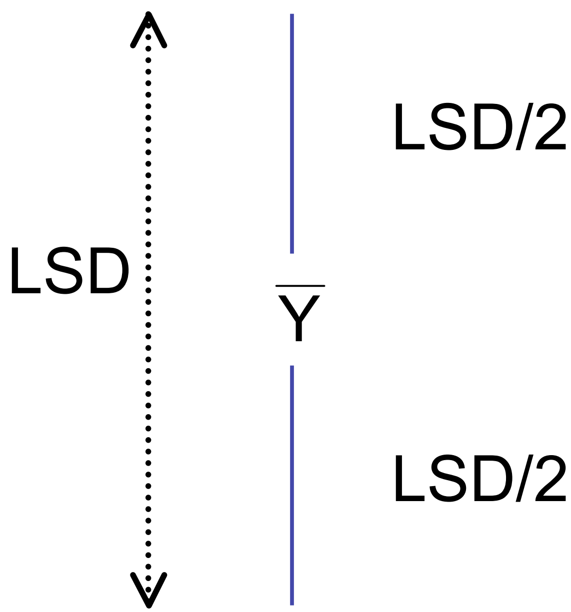
$$\frac{Z * SE_{\overline{Y}_1-\overline{Y}_2}}{2}$$

$$\overline{Y}1$$

Daniel Smith
California Dept of Health Services

E H I B

LSD = Least significant difference



Where LSD = Z*(SE of the difference
between two means)

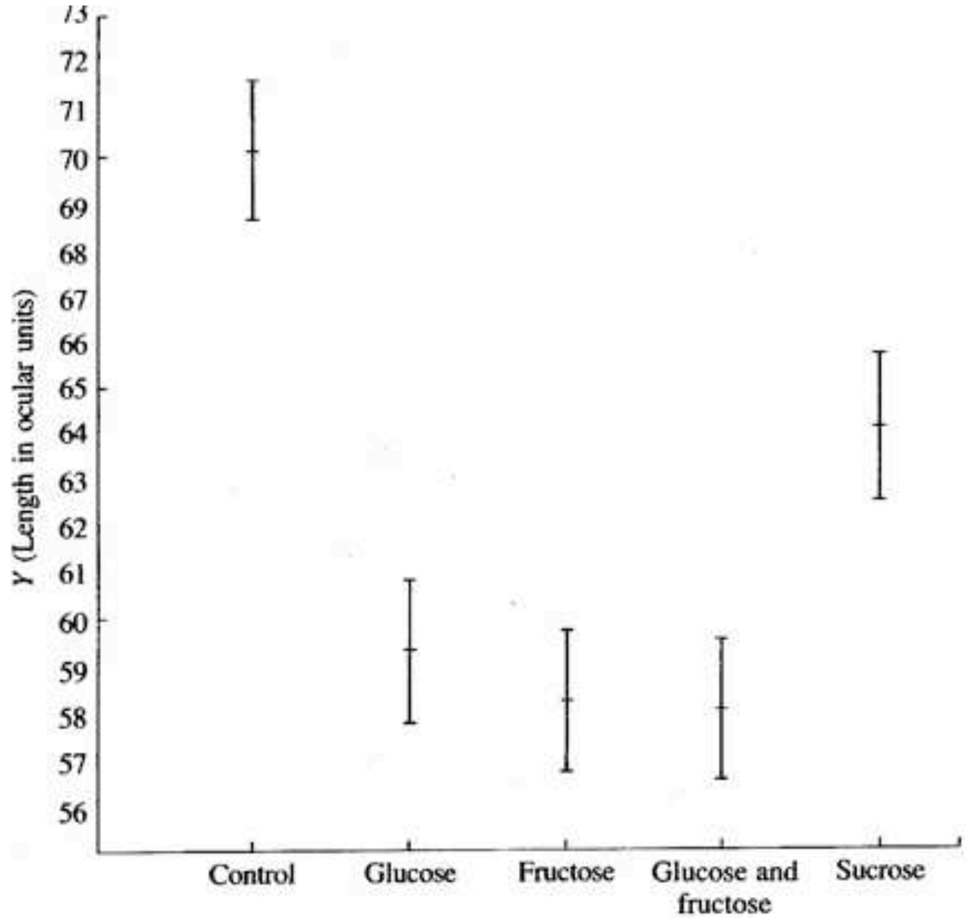
These are *not* confidence intervals,

But “error bars” that show the distance between two significantly different values

Uses a *pooled standard error*,
so can compare means across all groups

Daniel Smith
California Dept of Health Services

Example from Sokal and Rohlf (1995)



95% comparison intervals by the T-method for the means of the pea section data. Means whose intervals do not overlap are significantly different.

Daniel Smith
California Dept of Health Services

How to do this for epidemiologic measures?
First, consider Poisson rates…

$$\text{Poisson rate}: \quad \frac{\text{cases}}{\text{person-years}} = \frac{a}{N}$$

$$\text{Variance of } \sqrt{a} = \tfrac{1}{4}$$

$$\text{Variance of } \textit{difference } \sqrt{a}_1 - \sqrt{a}_2 = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}$$

$$SE = \sqrt{Var} = \sqrt{\tfrac{1}{2}}$$

E H I B

Daniel Smith
California Dept of Health Services

An LSD-style interval for Poisson rate can be constructed

$$\text{LSD} = Z\sqrt{\tfrac{1}{2}}$$

$$\text{Each arm of "LSD Interval"} = \frac{Z\sqrt{\tfrac{1}{2}}}{2}$$

$$\text{For } Z = 1.96, \text{arm} = 0.693$$

$$\mathit{95\%\ LSD\ limits} = \frac{\left(\sqrt{a} \pm 0.693\right)^2}{N}$$

EHIB

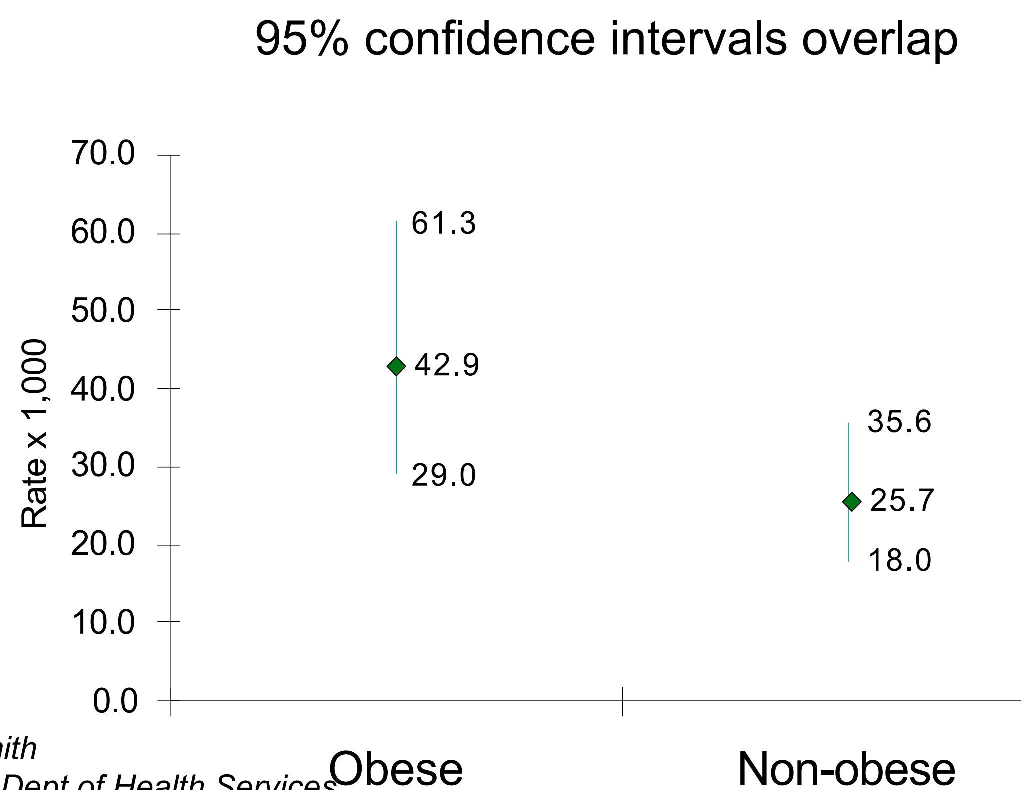
Daniel Smith
California Dept of Health Services

Example 1 again

	Obese	Non-obese
Deaths	30	36
Person-years	699	1399

χ^2 = 4.38
P value = 0.036

RR=1.67
95% CL: 1.03-2.71



Daniel Smith
California Dept of Health Services

E H I B

Example 1 again

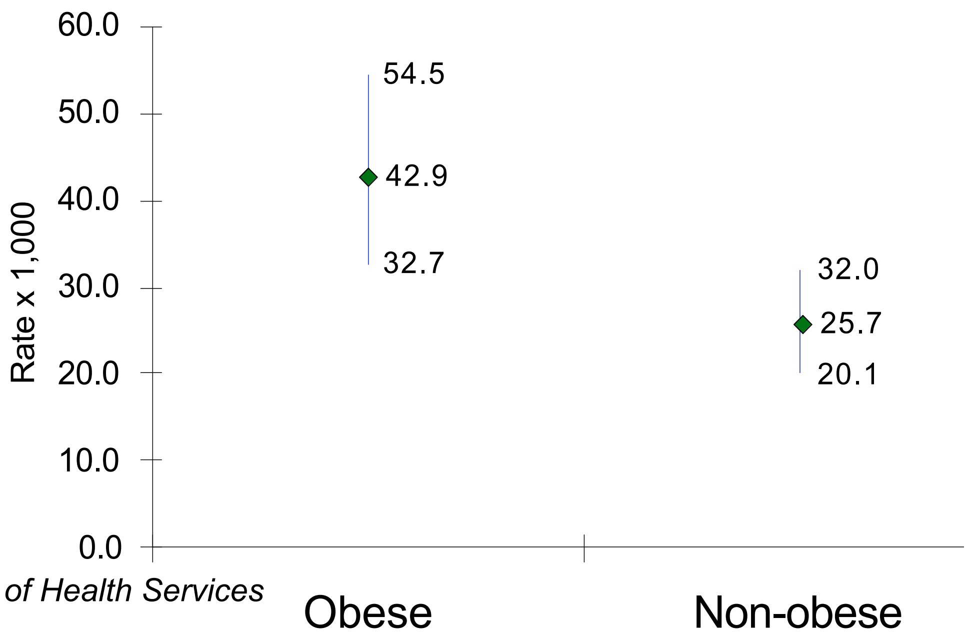
	Obese	Non-obese
Deaths	30	36
Person-years	699	1399

χ^2 = 4.38
P value = 0.036

RR=1.67
95% CL: 1.03-2.71

$$\textit{95\% LSD limits}: \frac{\left(\sqrt{30} \pm 0.693\right)^2}{699} = 32.7/1{,}000 \ - \ 54.5/1{,}000$$

$$\frac{\left(\sqrt{36} \pm 0.693\right)^2}{1399} = 20.1/1{,}000 \ - \ 32.0/1{,}000$$



95% LSD-type intervals don't overlap
(Consistent with χ^2 test)

Daniel Smith
California Dept of Health Services

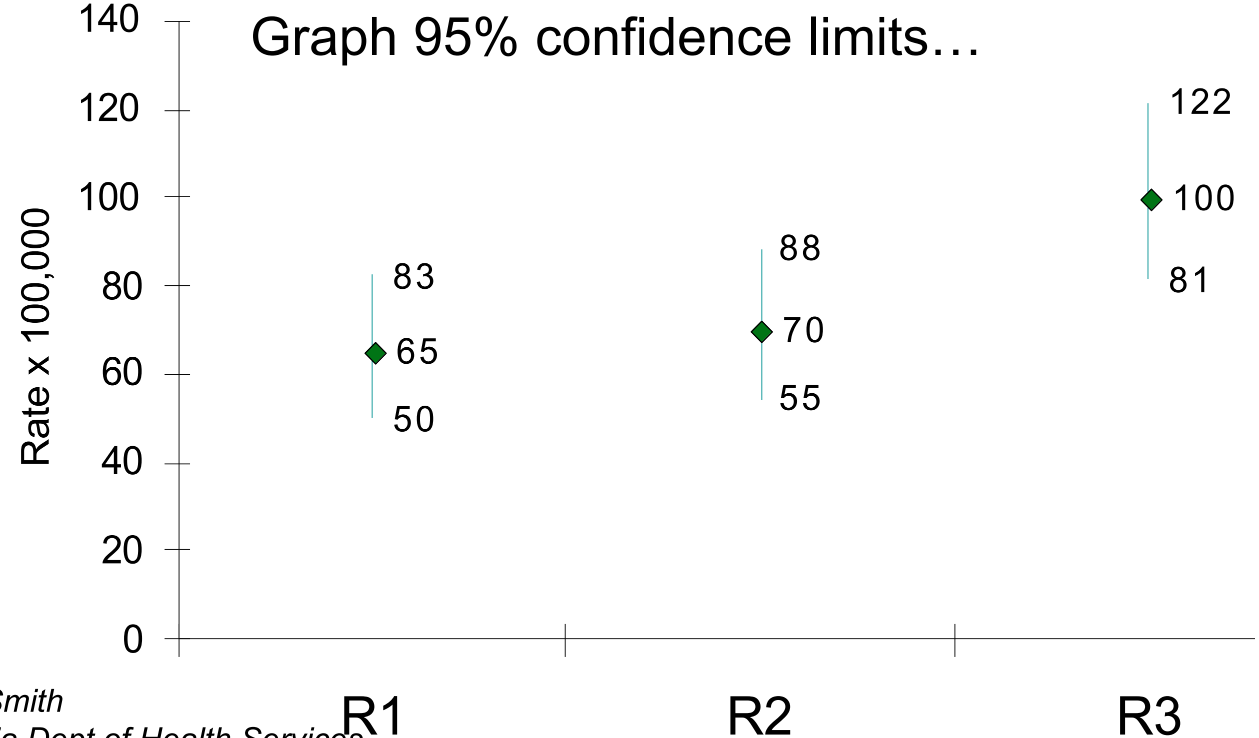
Three or more rates…

R_1	R_2	R_3
65/100,000	70/100,000	100/100,000

$P = 0.67$ (R_1 vs R_2) $P = 0.02$ (R_2 vs R_3)

$P = 0.006$ (R_1 vs R_3)

Graph 95% confidence limits…



All three confidence limits overlap!

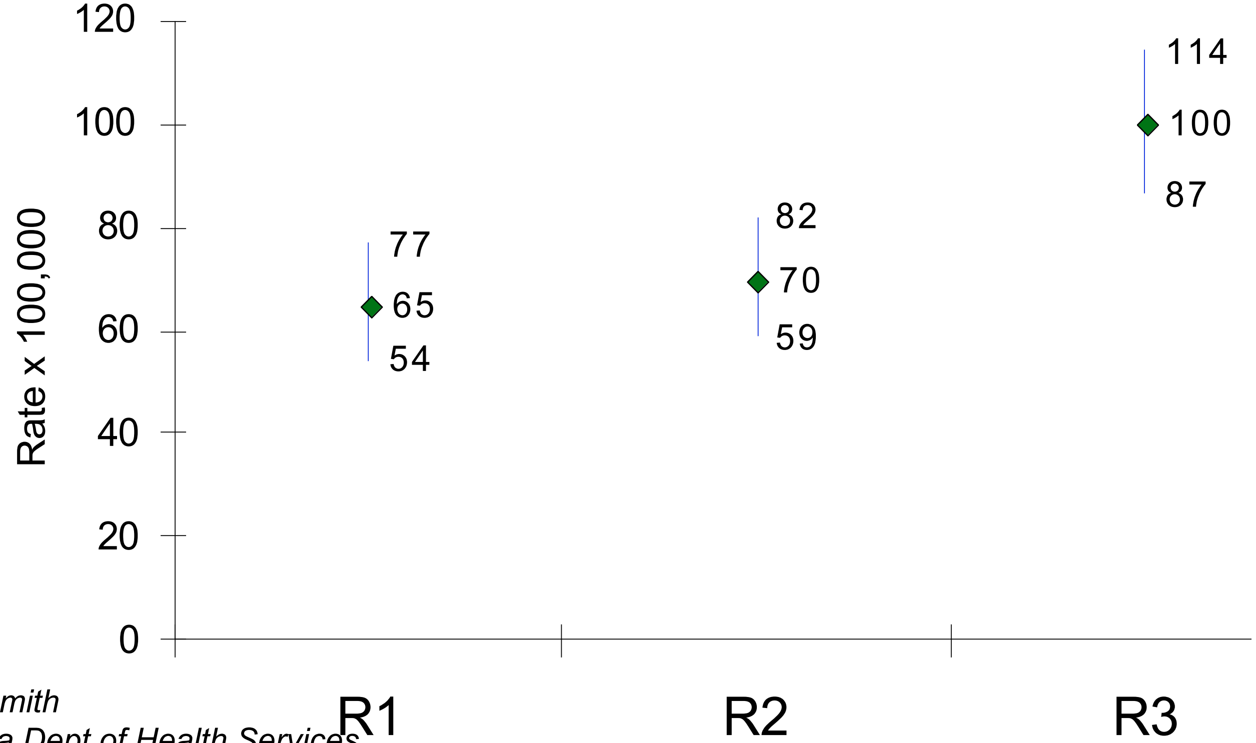
Daniel Smith
California Dept of Health Services

Three or more rates with LSD-style intervals

R_1	R_2	R_3
65/100,000	70/100,000	100/100,000

P = 0.67 (R_1 vs R_2) P = 0.02 (R_2 vs R_3)

P = 0.006 (R_1 vs R_3)



Now, overlaps agree with tests

Daniel Smith
California Dept of Health Services

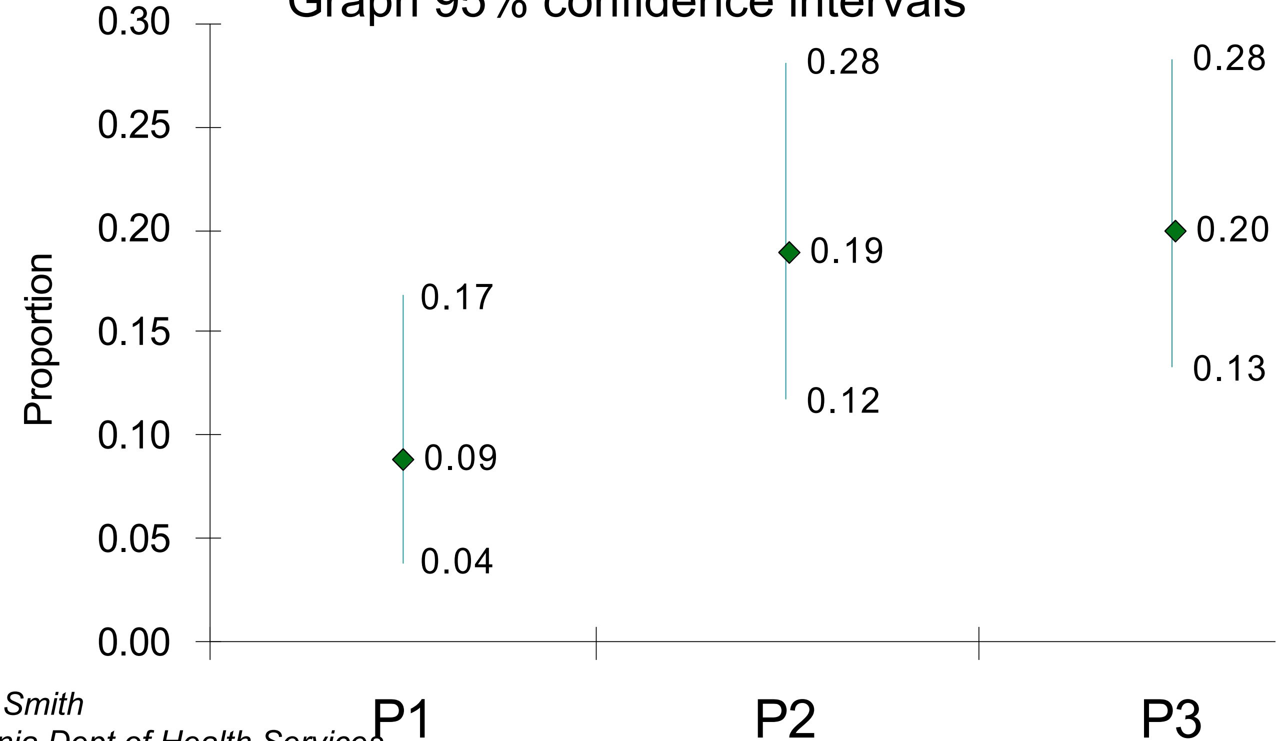
Try to adapt this to binomial case: Three or more proportions…

P_1	P_2	P_3
8/90	19/100	24/120

P = 0.047 *P* = 0.85

P = 0.027

Graph 95% confidence intervals



All three confidence intervals overlap!

Daniel Smith
California Dept of Health Services

Three or more proportions with LSD-style intervals

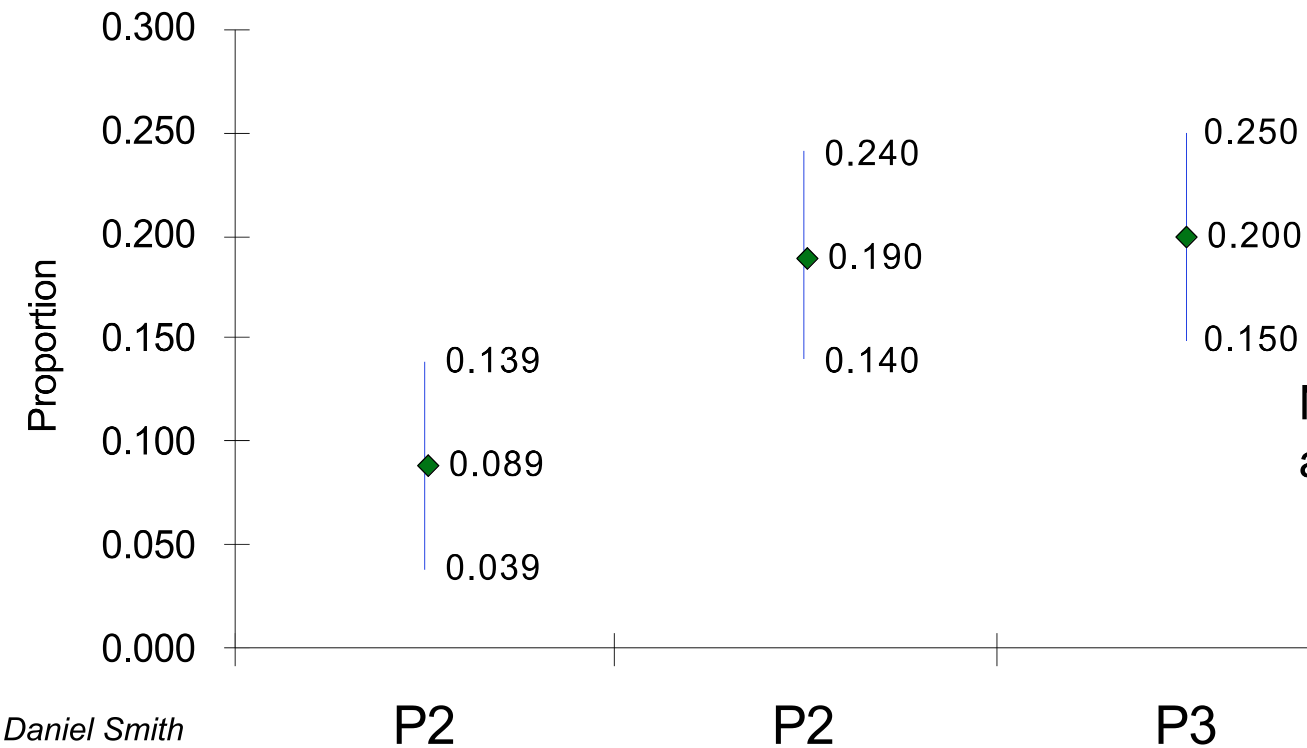
P_1 8/90 P_2 19/100 P_3 24/120

P = 0.047   *P* = 0.85

P = 0.027

Construct LSD-type intervals based on standard error of a test of two proportions:

$$\text{LSD Interval}: \; P \pm \frac{Z\sqrt{\frac{2\bar{p}\bar{q}}{\bar{n}}}}{2}$$



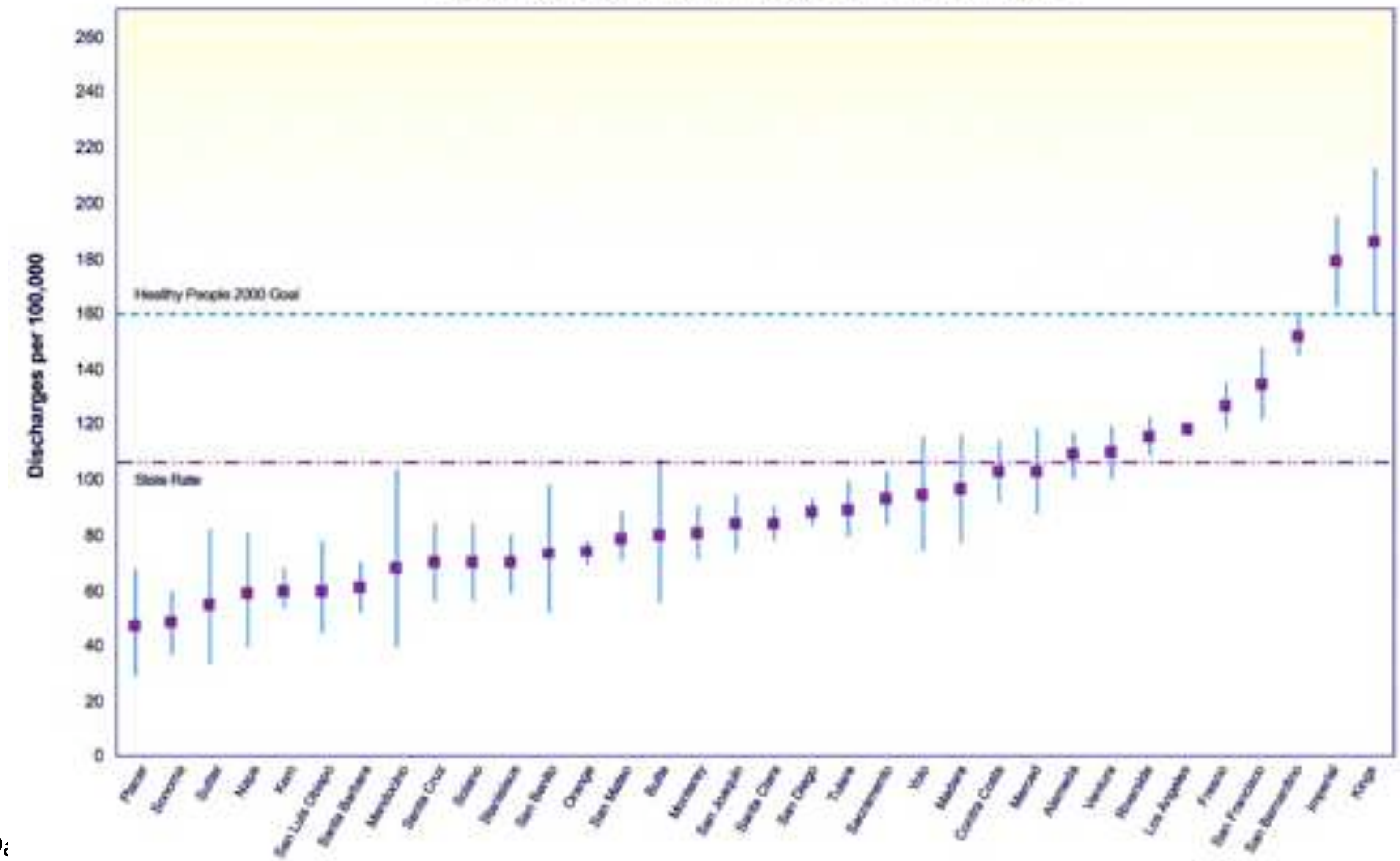
Now, overlaps agree with tests

EHIB

Daniel Smith
California Dept of Health Services

This *is* OK…
Comparing rates to a hypothetical value

Figure 4: Age-Adjusted* Asthma Hospital Discharge Rates for Hispanics by County, 1995-1997, with 95% Confidence Intervals.

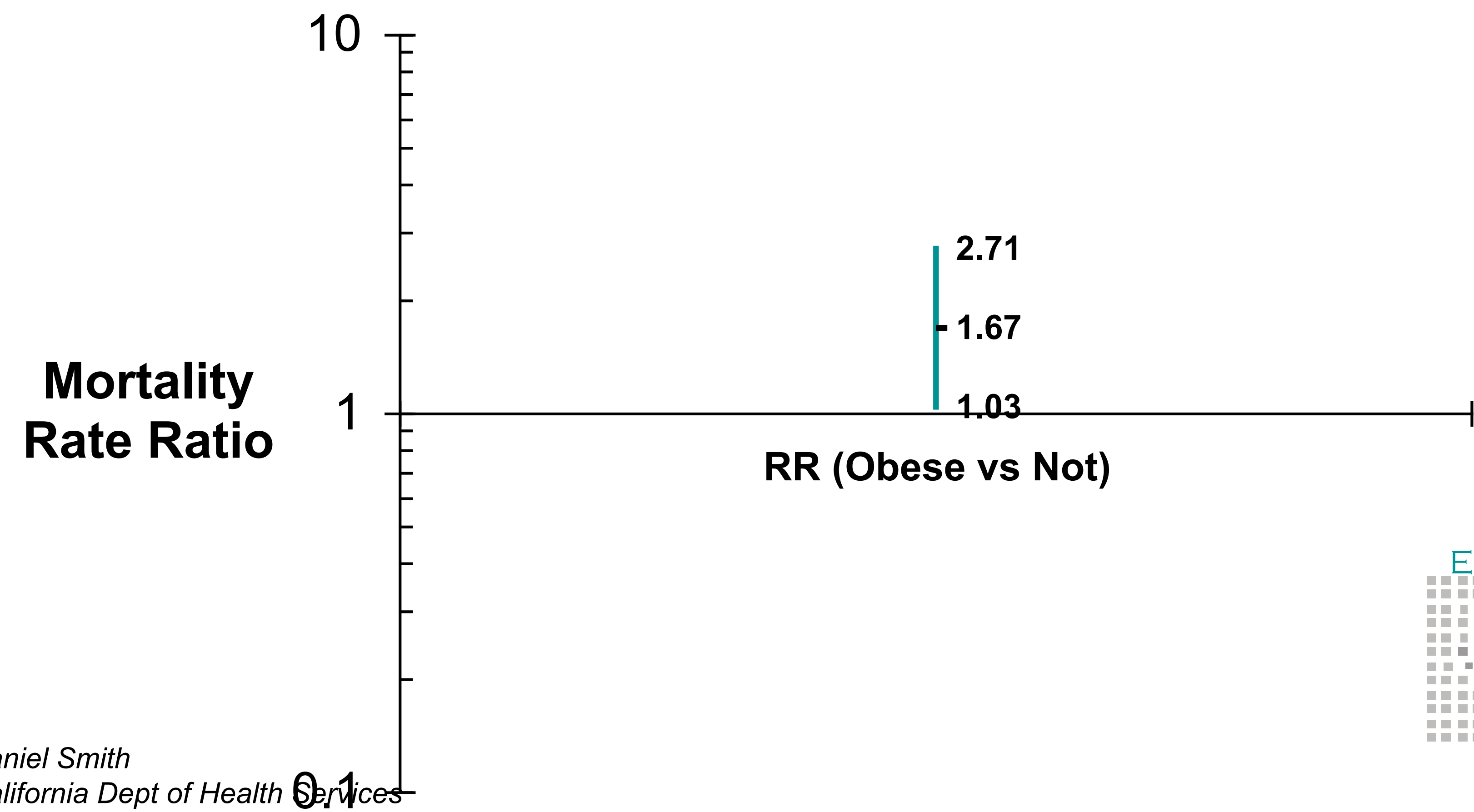


*Age-adjusted to the 1990 California population. Counties with less than 20 cases not shown.

Da
Ca

This is also OK…

Confidence limits for a *comparative* measure, versus the null hypothesis value

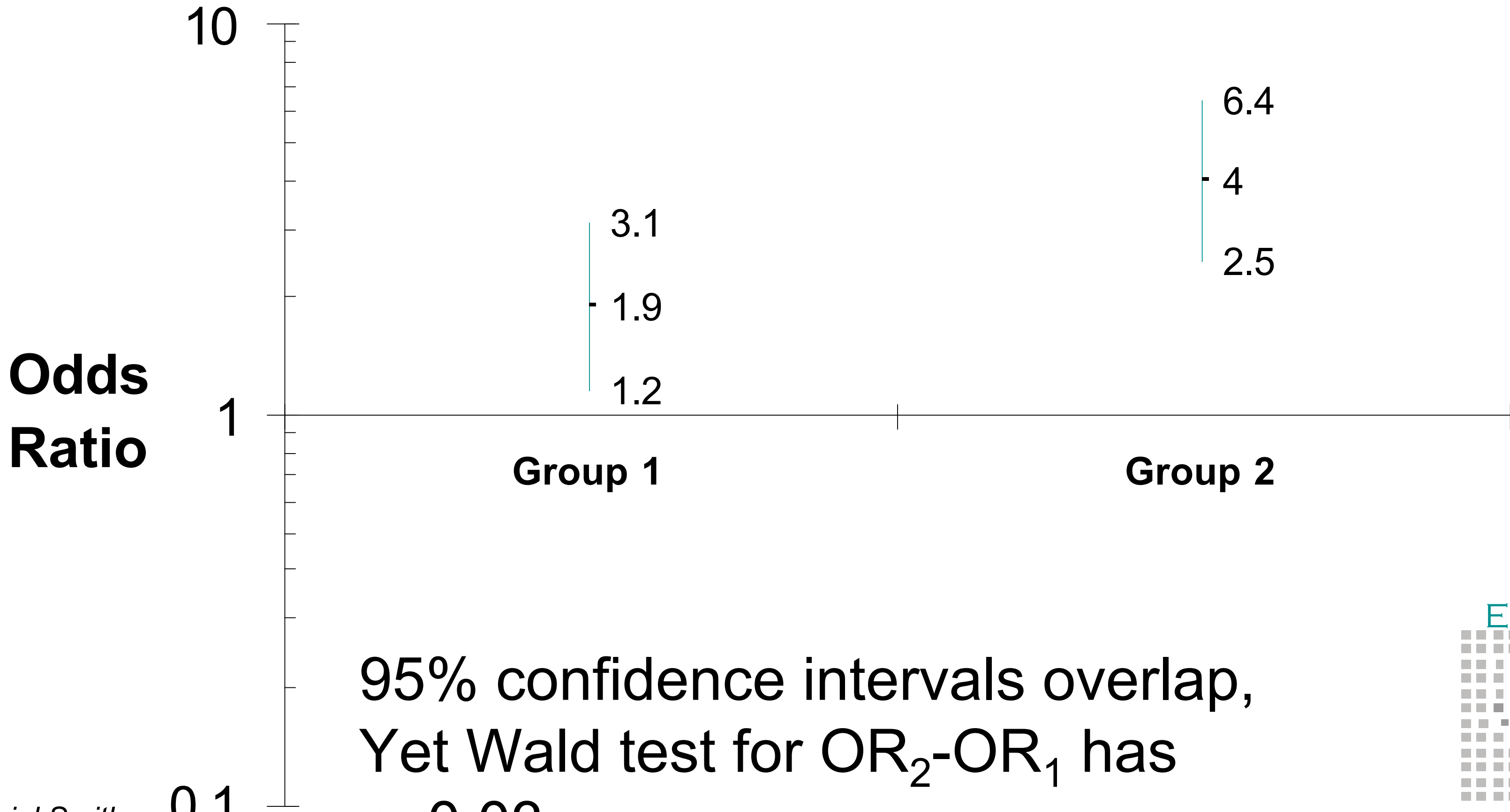


E H I B

Daniel Smith
California Dept of Health Services

But it is *not* OK to use intervals to compare OR_1 to OR_2…

Confidence interval around *one* OR is not a comparative measure between *two* ORs



95% confidence intervals overlap, Yet Wald test for OR_2-OR_1 has p=0.03

Daniel Smith
California Dept of Health Services

EHIB

So what to do?

- Be careful about use of overlap criterion:
  - –Overlap of 1-α % intervals means *p* could still be < α
  - –Intervals can overlap as much as 29% and rates can still be different at the 0.05 level (van Belle, 2001)
  - –Non-overlap of 1-α % confidence intervals means that *p* << α
- Show confidence intervals, but don't use them for testing between two groups
- Show LSD-type intervals, but don't call them confidence intervals

Epilogue

What to do for *adjusted* or *standardized* rates?
(Previous examples have been crude rates)

- It has yet to be worked out how to construct the LSD intervals for *weighted averages…*
- But provisionally, you might try this approach:
  - Note that, for crude rates, the LSD-type interval is narrower than the standard confidence interval by the factor of √1/2
  - Can create an LSD-type interval by multiplying standard CI by √1/2 :

    Width of interval = +/- Z*SE*(√1/2), where SE is the standard error of the standardized rate.

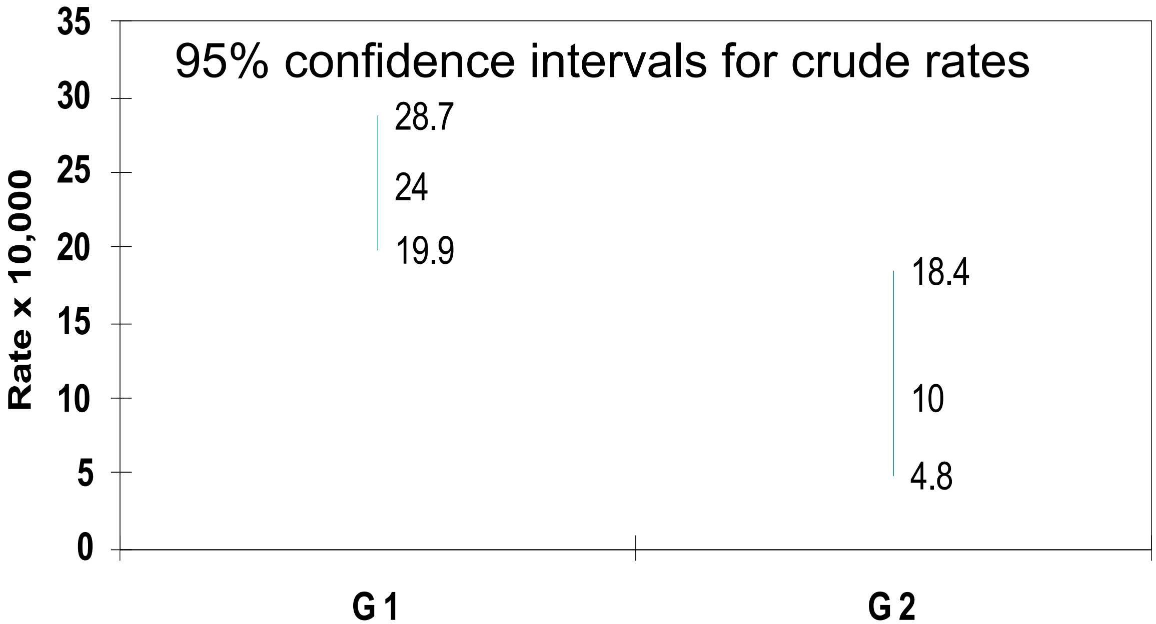
E H I B

Daniel Smith
California Dept of Health Services

Example of stratified data with confounding

	Group 1	Group 2	$RR_{stratum}$
Stratum 1	20/15,000	5/6,000	1.60
Stratum 2	100/35,000	5/4,000	2.29
Crude totals	120/50,000	10/10,000	2.40

Crude χ^2
p=0.006



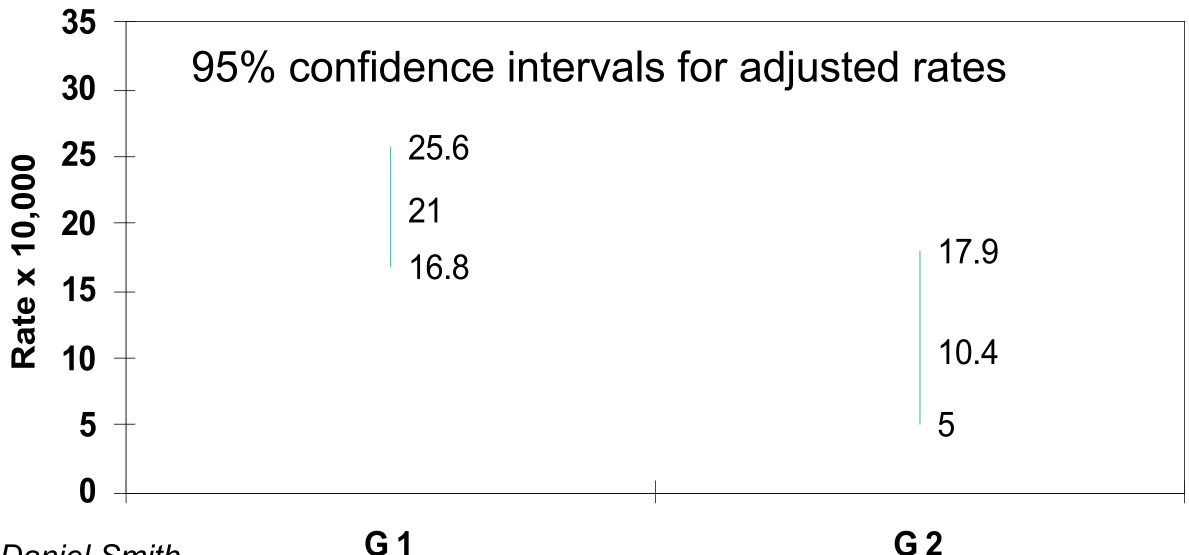
Confidence intervals don't overlap, but rates are confounded (crude RR is greater than either stratum RRs)

Daniel Smith
California Dept of Health Services

Standardized rates
(“direct adjustment,” using weights)

	Group 1	Group 2	Weight
Stratum 1	20/15,000	5/6,000	0.5
Stratum 2	100/35,000	5/4,000	0.5
Standardized rate	21.0/10,000	10.4/10,000	

Standardized RR = 2.01 (95% CL 1.02 – 4.24)
Mantel-Haenszel χ^2 p-value = 0.039

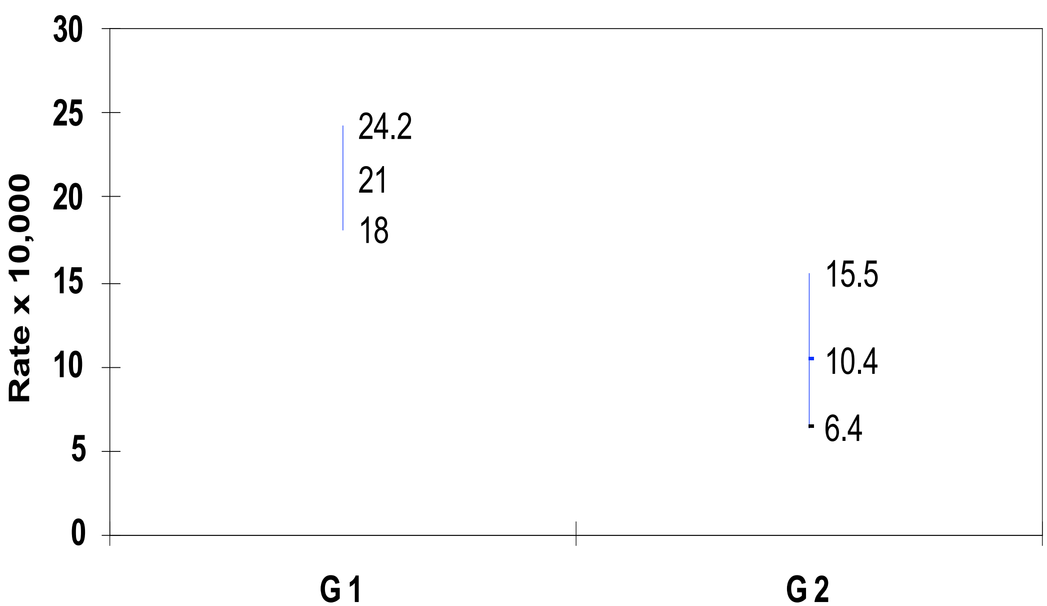


As expected, the 95% confidence intervals for the standardized rates inappropriately overlap

Daniel Smith
California Dept of Health Services

Try the correction factor…

Calculate intervals using $Z * SE_{(\text{standardized rate})} * \sqrt{1/2}$



- Error bars don’t overlap
- Now they’re consistent with Mantel-Haenszel χ^2 test controlling for stratum

Daniel Smith
California Dept of Health Services

References

- LSD Intervals for ANOVA
  – Andrews HP, Snee RD, Sarner MH. Graphical display of means. *American Statistician* 1980;34:195-199
  – Snedecor G, Cochran G. *Statistical Methods*, 8th Ed. Ames: Iowa State University Press, 1989; 235-236
  – Sokal RR, Rohlf FJ. *Biometry,* 3rd Ed. New York: WH Freeman, 1995; 243-246
- Poisson properties
  – Armitage P, Berry G. *Statistical Methods in Medical Research*, 2nd Ed. Oxford: Blackwell, 1987; 362-363
- Overlap issues in general
  – Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician* 2001; 55(3):182-186
  – van Belle G. *Statistical Rules of Thumb*. New York: Wiley, 2002; 39-40

E H I B