

Proper Use of Overlap Between Confidence Intervals to Examine Differences

Gleb Haynatzki, PhD

Biostatistics Department

University of Nebraska Medical Center



Learning Objectives

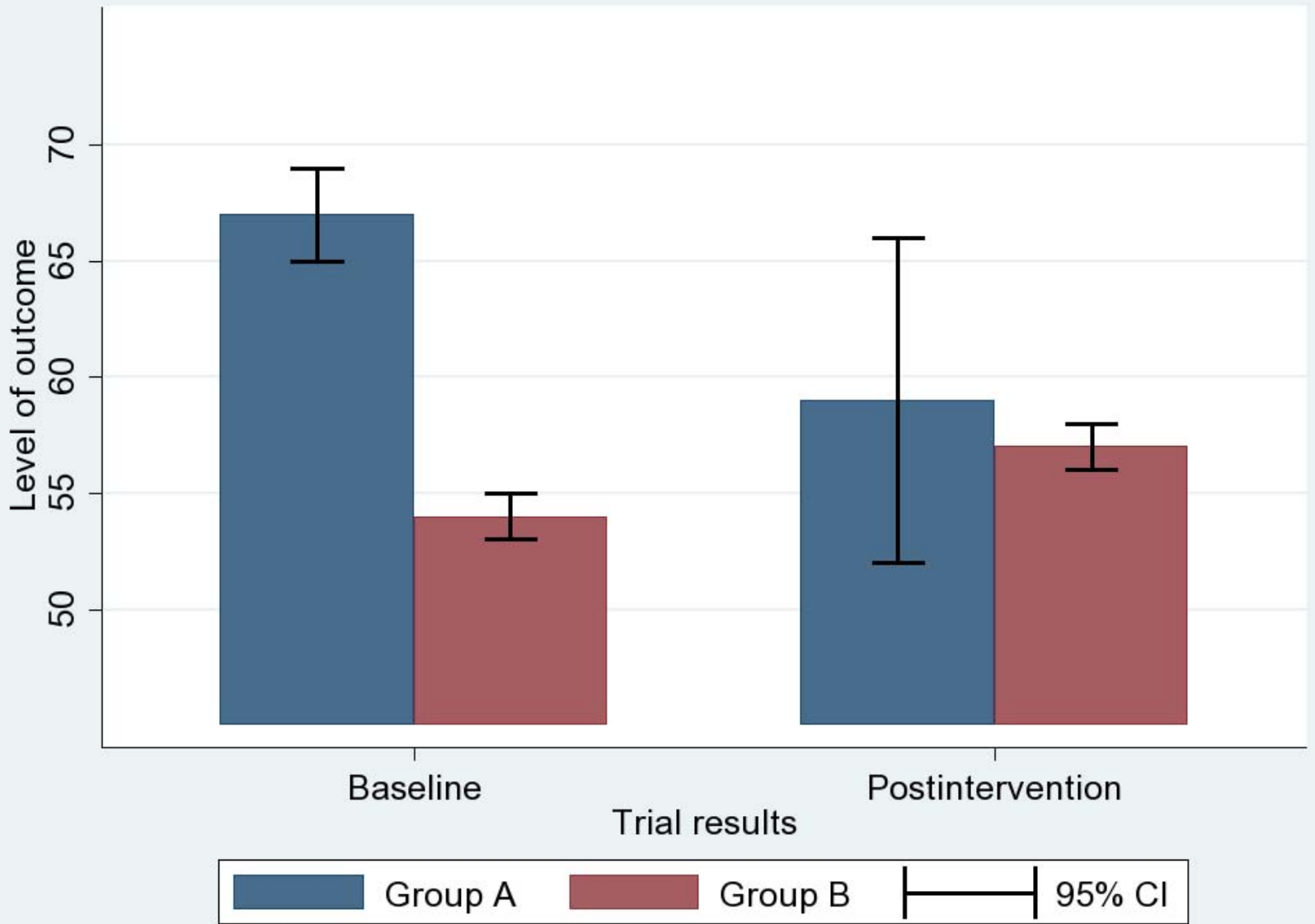
- Problem formulation



Abstract (1): Problem formulation

- To judge whether the difference between two point estimates is statistically significant, data analysts often examine the overlap between the two associated confidence intervals
- This method is simple because it is easy to compare boundary values from two confidence intervals to see whether the intervals overlap
- Moreover, it seems natural to inspect graphed confidence intervals for overlap
- Q: The method implies that 95% CIs just touching end to end is equivalent to statistical significance, $p < .05$, but is it true? In fact, often the p-values are much smaller than .05
- Examples of bar graphs with 95% CIs are shown in the next 3 slides

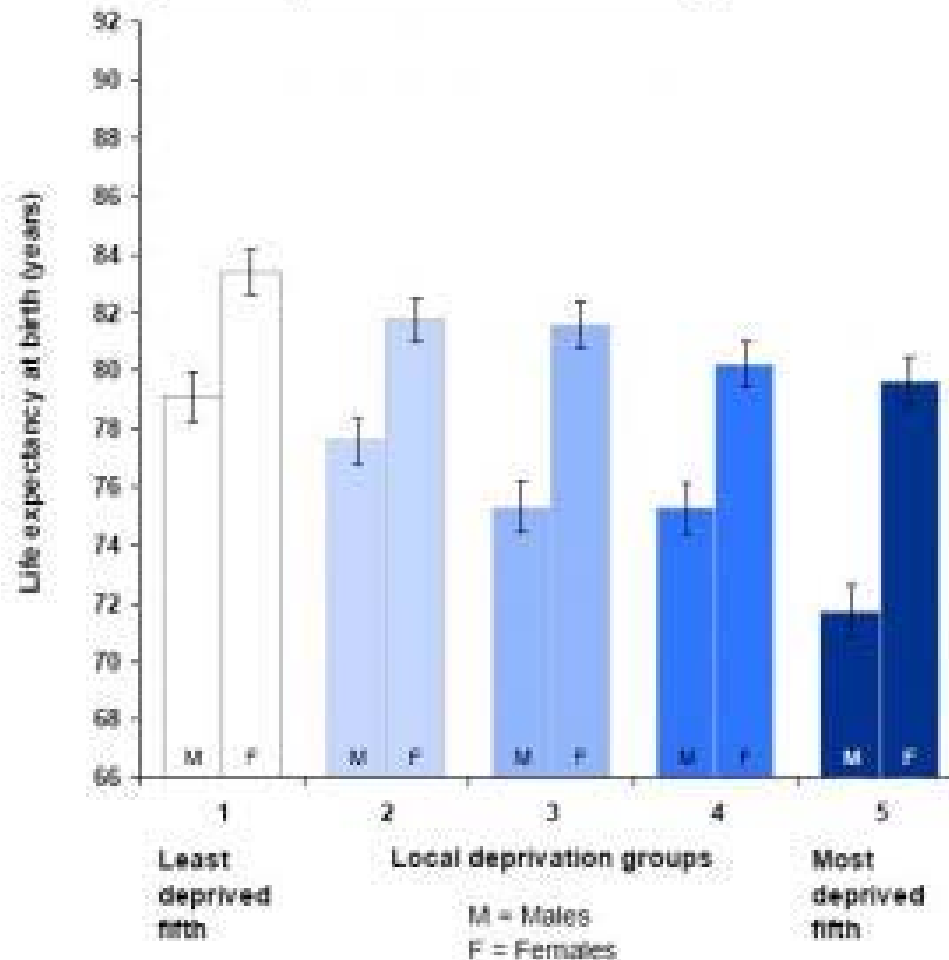
Ex. 1



Health inequalities: a local perspective

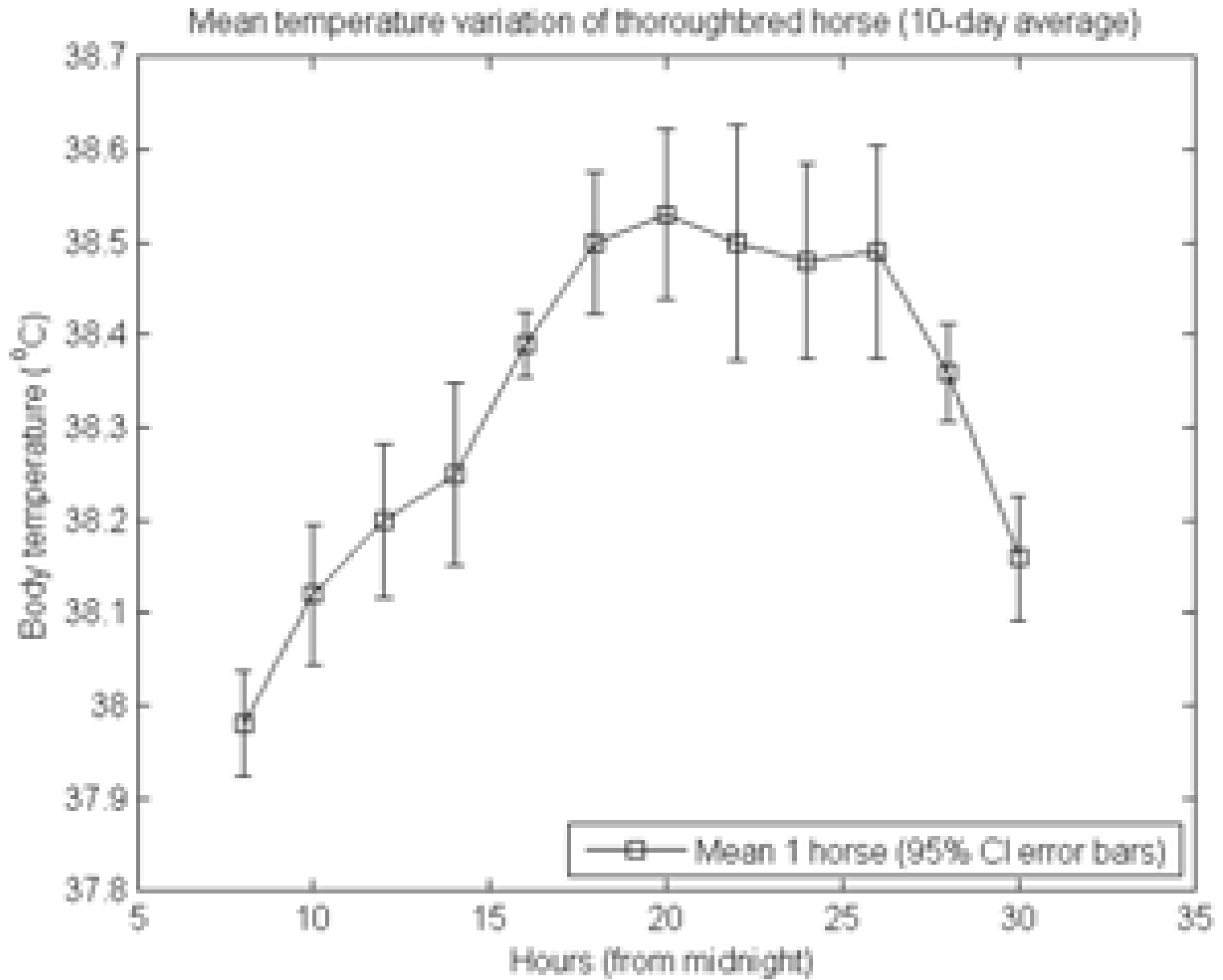
Ex. 2

Inequalities in life expectancy (2002-2006) for men and women who live in areas with different levels of deprivation (within this local authority).



I 95% confidence interval. These indicate the level of uncertainty about each value on the graph. Longer/wider intervals mean more uncertainty. When two intervals do not overlap it is reasonably certain that the two groups are truly different.

Ex. 3





Abstract (2): What will be done

- We will compare the method of examining overlap to the standard method of *testing significance* under the common assumptions of (i) consistency, (ii) asymptotic normality, and (iii) asymptotic independence of the estimates
- That comparison will be done on a base (but important) case and demonstrated in a figure



Abstract (3): Summary of conclusions (1)

- Rejection of H_0 by the method of examining overlap implies rejection by the standard method
- Failure to reject H_0 by the method of examining overlap does **not** imply failure to reject by the standard method
- Thus:
 - the method of examining overlap is slightly more conservative (ie, rejects H_0 less often) than the standard method when H_0 is true, and
 - the method of examining overlap mistakenly fails to reject the H_0 more frequently than does the standard method when H_0 is false



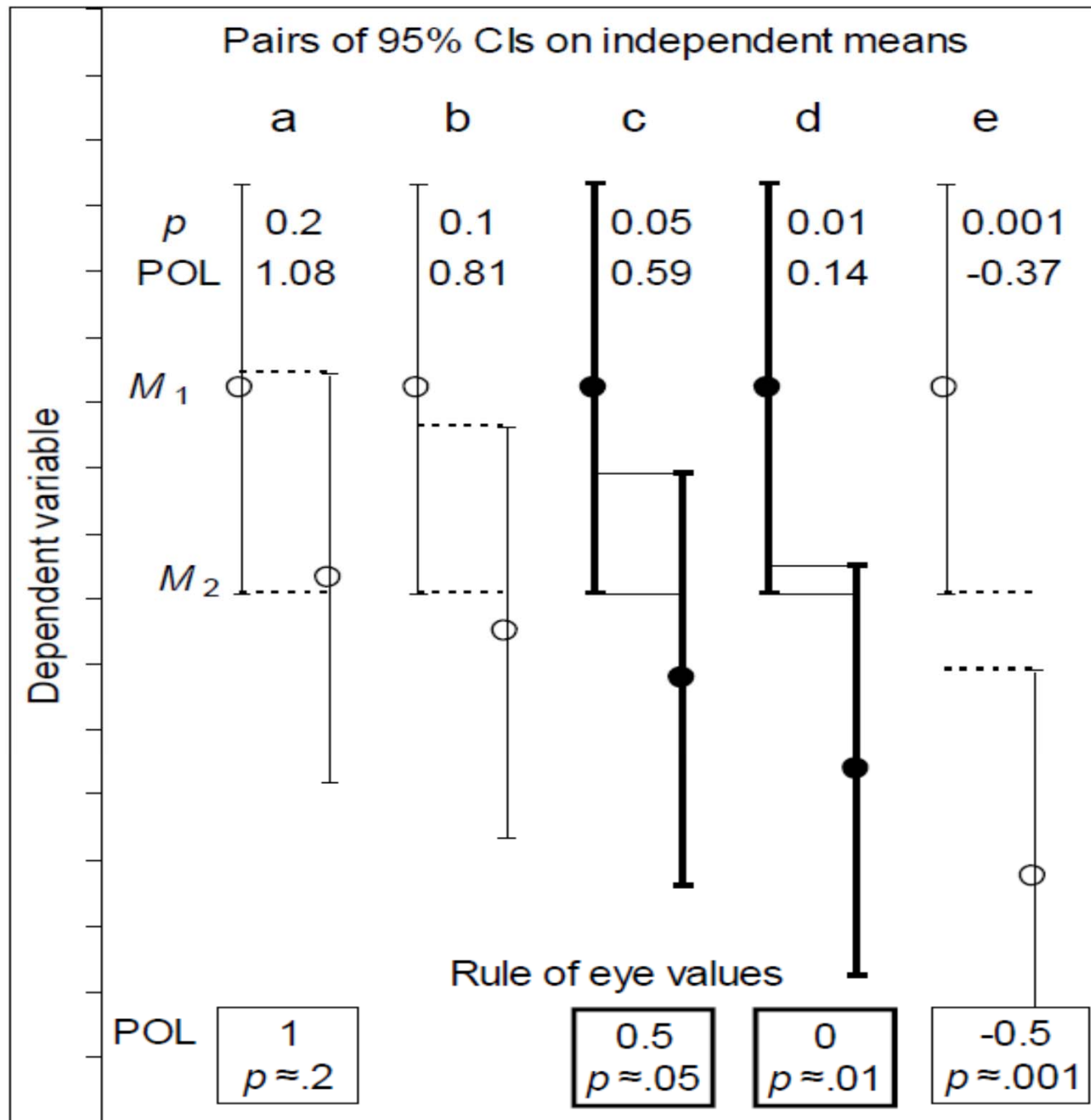
Abstract (4): Summary of conclusions (2)

- Although the method of examining overlap is simple and especially convenient when lists or graphs of confidence intervals have been presented, we conclude that it should **not** be used for formal significance testing unless the data analyst is aware of its deficiencies and unless the information needed to carry out a more appropriate procedure is unavailable



Base case

- Consider the next figure that shows 95% CIs on the means of *two* samples that are
 - independent
 - of equal sample sizes n
 - where n is “large”
 - from normal populations
 - with equal variances
- Several subcases (**a-e**) are considered





Base case: Comments (1)

- Means M_1 and M_2 , and 95% CIs for pairs of indep samples, with equal n , from normal populations with equal variances
- The CIs are calculated using z , so it is assumed n large
- **Proportion overlap (POL)** is overlap expressed as a proportion of the length of a single arm of a CI (ie, the margin of error)
- A gap between CIs is signaled by a negative POL value (Pair **e**)
- M_2 is varied to achieve selected values of 2-tailed p
- As POL decreases, the p value decreases (shown near the top)



Base case: Comments (2)

- ***The rule of eye*** simply states that $\text{POL} = 0.50$ implies $p \leq .05$
- *The rule of eye* (p , POL) pairs in the lower boxes provide approximate benchmarks for estimating the p value for any *observed* amount of overlap of two independent 95% CIs
- *The rule of eye* is slightly conservative, eg, Pair **c** shows that $\text{POL} = 0.59$ gives $p = .05$, but *the rule of eye* specifies a POL of approximately 0.50 or less, for $p \leq .05$
- The relationships are practical when $n \geq 10$, and the margins of error do not differ by more than a factor of 2.



Literature

- Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, 28, 205-220.
- Schenker, N, Gentleman, J. F. (2001). On judging significance of differences by examining the overlap between confidence intervals, *The American Statistician*, 55, 182-186.
- Cole, S. R., and Blair, R. C. (1999), "Overlapping Confidence Intervals," *Journal of the American Academy of Dermatology*, 41, 1051-1052.



Thank you!