

Heisler M, Faul JD, Hayward RA, Langa KM, Blaum C, Weir D. Mechanisms for Racial and Ethnic Disparities in Glycemic Control in Middle-aged and Older Americans in the Health and Retirement Study. *Arch Intern Med* 2007; **167**: 1853-1860.

This study investigated differences in HbA1c among self-characterized diabetics who were white, black, or Latino. It found worse mean values among black and Latino respondents than among white respondents, with little of the differences explainable with other measured variables.

**Problem 1 (negative conclusions based on p-values).** The initial result reported in the Abstract was: “There were no significant racial/ethnic differences in HbA1c levels in respondents not taking antihyperglycemic medications.” This appears to be based on the p-values in Table 1. The actual black-white difference in mean HbA1c was  $7.17-6.39=0.78$ ; this difference was not actually given in the paper, so of course no CI for it was given either. We can deduce the CI’s based on the p-values given in Table 1: for not on medication we have 0.78 (-0.36 to 1.92) and for on medication we have 0.85 (0.34 to 1.36). This is a lot of overlap, and we can assume that the difference of 0.78 is large enough to be important because it is nearly as large the difference they do investigate (0.78, versus a 0.85 difference among those on medication) and actually larger than the adjusted difference (0.73) from their final multivariate model. For the black-white disparity, the main difference between those on vs off medication was therefore the larger p-value due to the small number of subjects not on medication. For Latino respondents, the contrast between those on and off medication is larger. We can deduce a p-value for the interaction of Latino vs white with medication status from the information they give, obtaining  $p=0.17$ . Thus, while there is a suggestion of a different white-Latino disparity by medication status, the evidence is not conclusive.

Fortunately, the authors did not attempt to draw any conclusions from the “lack” of a disparity among those not taking medication. But this does not mean that others will not mistakenly attempt to do so.

They did note that multiple regression models that included all respondents produced similar results to the medication-only analyses. Their main conclusion in the Abstract does not distinguish those on and off medication.

**Problem 2 (vague, misleading phrasing).** In the discussion of limitations on p. 1859, they write concerning inaccurate self reporting and social desirability bias, “There is no evidence to suggest, however, that different racial/ethnic groups are more susceptible to such bias than others.” We cannot tell whether this has been studied and racial/ethnic differences are known to be small, or if there is simply nothing known about the issue. Also, the blanket claim that “there is no evidence” would require an extensive literature review, and the lack of any citations for this statement casts doubt on whether any such review was done. It probably would have been better for them to omit this sentence.

**Best Practice 1 (provide directly relevant estimates and CI’s).** Table 3 shows the estimates of primary interest, but without CI’s. (The only CI’s given in the entire paper are in the footnote to Table 1, and those only address overall means of HbA1c.) The CI’s for the racial/ethnic differences in Table 3 cannot be deduced, because exact p-values are not provided.

**Best Practice 2, 2a (interpretation based on estimates and CI’s).** See Problem 1. They do interpret the quantitative estimates of how much of the disparities were explained by the measured covariates (14% and 19%, given at the top of the right column on p. 1856), using the phrase “only partially explained”. The results and interpretation suffer from lack of CI’s to show the strength of evidence for the impression that little was explained. Obtaining these would probably require bootstrapping, but would not be especially difficult.

**Best Practice 3 (discuss what may be true in general).** There was some sloppy use of “significant” alone (see below) that could lead to confusion on this issue and which related to their difficulties with Problem 1.

**Best Practice 4 (state what you did find or learn).** Mostly OK, but see Problems 1 and 2.

**Best Practices 5, 5a (learn as much as you can).** Table 3 shows a thorough examination of their main question, what may explain disparities, although they could have learned more by examining CI’s, notably for how much was explained. They did some sensitivity analyses to confirm their main findings.

**Best Practices 6, 6a (include scientific considerations).** The design, analysis, and interpretation appear to have been informed by existing theories and knowledge about the issue under study (see their Figure). They did not use multiple comparisons adjustments.

**Best Practice 7 (prefer accuracy to conservatism).** They do not appear to have introduced conservative bias in the main analyses.

**Checking assumptions.** They report performing some regression diagnostics.

**Missing data.** They attempted to avoid problems due to missing data, but it appears that they used only single imputation instead of more rigorous multiple imputation methods. They also overstated what they found as “no differences in results”, instead of “no qualitative differences in results”.

**Insufficient precision.** Some p-values reported as “0.02” could be anywhere in the range 0.015 to 0.025, which seems undesirably wide.

**Use of “significant” alone.** This is frequent in this paper, with the meaning sometimes being “statistically significant” and sometimes being “important”. For example, they describe the results of some sensitivity analyses as, “Results did not differ significantly using these cutoff values, so we report only the linear regression model results.” They also note “significantly greater racial/ethnic disparities in respondents younger than 65 years” (p. 1858). This apparently does not mean that they tested the interaction of age with racial/ethnic category and found  $p < 0.05$ , and we cannot tell from the (skimpy) information given whether such a test might have produced  $p < 0.05$ . Note also that the black-white disparity among those not on medication is larger for those  $\geq 65$  (Table 1).