Reviewer's report

Title: Health Plan Administrative Records versus Birth Certificate Records: Quality of Race and Ethnicity Information in children

Version: 1 Date: 17 August 2010

Reviewer: Karl Eschbach

Reviewer's report:

This study investigates the accuracy of administrative records reporting of race and Hispanic origin for children, with the birth/death certificate treated as the standard in this analysis. This is an interesting exercise, that adds to the literature. The use of birth certificate as the standard seems appropriate in comparison, especially insofar as official uses of data in the federal statistical system use birth certificate classifications for assigning race ethnicity to new-born children in intercensal population estimates, until census self (parent)-identification recalibrates. Thus denominators used, for example, by NCHS, SEER and others rely on this standard. At the authors discretion, this might be added as an additional justification of this source.

I have two related questions about the birth certificate standard:

1) Does the KPSC distribution precisely mirror what is on the birth certificate, or does it approximate it? Thus, were missing cases on the birth certificate actually missing on the birth certificate, or were they missing in KPSC transcription of what was on the birth certificate.

2) Are we sure that persons who fill out administrative records do not consult the birth certificate database to assign race/ethnicity. I assume the answer is "yes." Because the database is itself an in-house administrative file at KPSC, it would be useful if the authors used half a sentence to assure the readers of this point.

It would also help if the authors include a discussion of the decisions they make about how to code the information from the birth certificate database. In particular, on page 6 it is stated that where mother and father were of different race, that the standard was set to "multiple races." This is reasonable, but are there other resolutions? Specifically, the authors may want to consider referencing the practices used in the federal statistical system and discussed in Vital Health Statistics Series 2, Number 135: http://wonder.cdc.gov/wonder/help/populations/bridged-race/VitalHealthStatistics-Series2No135.pdf. The federal practices are not necessarily a gold standard, but they are important.

The federal standard laid out in the revised OMB statistical directive #15 privileges self-identification as the standard for assigning race/ethnicity. We know that in self-identification, respondents often simplify multiple ancestries. (Harris, David R. and Jeremiah Joseph Sim. 2002. "Who is Multiracial? Assessing the
In this case, failure to be concordant with assignment of multiple races would not necessarily represent erroneous report in administrative records. The data do not permit resolving who is simplifying in the administrative records (parent or administrative clerk), but at least this issue could be referenced and discussed.

The brief discussion of Hispanic reporting issues could be expanded. In many contexts, the concept of "Hispanic White" is not useful because most lay users of the terms assume that Hispanics and Whites are distinct groups. In southwestern states, racial statistics for "whites" that also include non-Hispanic whites, would confuse many people. Thus, it would be helpful to be absolutely clear about the PPV and sensitivity for the non-Hispanic racial categories, possibly by reproducing Table 2 data for Hispanic/non-Hispanic respondents separately, but at least by an expanded discussion of this stratification of the data.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.