Reviewer’s report

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

**Version:** 1 **Date:** 1 June 2010

**Reviewer:** Scarlett Gomez

**Reviewer’s report:**

Major compulsory revisions:

Were information on race and/or ethnicity available over time in the Kaiser EMR sources? If so, how was conflicting information derived from a single source, over time, resolved? Is it worthwhile to compare the quality (completeness and accuracy) of data across the multiple sources to determine whether one source is collecting more complete and/or accurate data than others? Instead of assuming that conflicting information across multiple EMR sources meant that the person was multiple race, an alternative strategy might be to determine whether one EMR source might be more accurate than others compared to birth certification information. That is, the information in the EMR source may conflict because one is correct, and one is not; with both maternal and paternal information available in the birth certificates, there is the opportunity to determine whether conflicting information means true multiple-race, or inaccuracies in one source.

How could there be a nearly null effect (OR = 1.01) for total number of medical encounters, but significantly higher effects for encounters in specific settings (inpatient, emergency), unless there were were also lower effects seen in other settings? It’s also not clear how there could be an effect for these types of encounters, if the ranges are so small, according to Table 1.

Since the general objective of this paper is to assess the extent to which EMR data on race and ethnicity among children has utility for research purposes, it would be able useful to examine whether data quality (completeness and accuracy) has changed over this 11-year period. Has data quality improved, worsened, stayed the same?

Minor essential revisions:

In the Discussion section, unless there were data that were not presented on changes in data quality over time, I don’t think that you can make this claim based on the findings that were presented: “The present study showed that the quality of information increased over time with increasing number of medical encounters, especially inpatient visits.”

Discretionary revisions:

Also from a utility standpoint, it might be useful to at least present the
distributions for Asian and Pacific Islander subgroups. In general, what is the level of coding detail that is available in the Kaiser EMR data?

Presumably the Kaiser EHR data systems did not allow for coding of multiple race (?). Has this practice changed?

It would be useful to present the % numbers in Figure 1.

**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.