Reviewer's report

Title: Transition to the New Data Collection Standards in the Department of Veterans Affairs

Version: 1 Date: 8 May 2006

Reviewer: Nancy Kressin

Reviewer's report:

General

This paper extensively examines the characteristics and quality of the new data VA is collecting on race and ethnicity. The authors have conducted a thoughtful and detailed analysis of the agreement of the self-reported race data with the prior (observed) variables on race, and have examined data accuracy within subgroups as well. I had relatively few criticisms as I thought the paper was clear and well written, and the conclusions directly derived from the results presented. Below, I list the comments I had.

-----------------------------------------------------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

According to my calculations, only about 20% of the VA population of interest (all VA healthcare users in 2004) were included in this study (52% of the individuals had a single self reported race value in 2004 could be linked to the older race data x 39% who had valid self-reported race) = ~20%. If this is correct, and only ~ 20% of the population were represented, discussion of the representativeness of this sample and generalizability of the findings should be included in the paper. Otherwise, if I misunderstand, it would be helpful to spell out more clearly the proportio of the parent population who were included in this study sample. The authors briefly touch on the related issue of the representativeness of the sample in the limitations paragraph of the paper, but I would encourage them to expand this discussion further.

-----------------------------------------------------------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

I would encourage the authors to comment on the finding that agreement rates varied by geographic region (and likely, by extension, on the racial/ethnic composition of geographic regions). That is, the rates were higher in the Midwest where there might be more whites, than in the West, where there might be more Native Americans or Hispanics.

I would also encourage the authors to clarify whether the fact that the parent study was concerned with urologic diseases is relevant to this paper. I was confused at first that the parent population was urology patients until I read the Study Design section several times to understand that this did not seem to be the case.

-----------------------------------------------------------------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)

What next?: Accept after minor essential revisions

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

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