Editor:

We thank both reviewers for their thorough reading and thoughtful comments on our paper. We present each reviewer’s comment verbatim followed by our response and action taken.

REVIEWER 1

Major Compulsory Revisions
The authors have a similar manuscript in press dealing with mammography. It should be very clear whether the mammography data reported in the present manuscript is the same data or whether new/different information is contained.

The in-press manuscript is attached. Although we used the same data for this paper, the analytic method is different. In the mammography paper, we assessed whether rates of self-report and verified claims (self-reported mammography confirmed by a mammography claim) differed. In this study, we examine rates of over-reporting (self-report in the absence of a claim) and under-reporting (claims in the absence of self-reports). Furthermore, the primary focus of the paper is the inclusion of five other preventive services. In other words, whereas the mammography paper focuses on just one preventive procedure, the current paper explores the extent to which discrepancies between self-report and claims may be a generalizable issue. Because the mammography sample size is larger than many of the other samples used in this paper, we aggregate across minority group whereas in the mammography paper we report results separately for African Americans and Hispanics. Thus, the overlap between the two papers is minimal.

Minor Essential Revisions
The first paragraph of Background states that there are differences in the size of the disparity for self-report and claims data with mammography. However, no reference is provided here.

References for this assertion have been provided.

It would be helpful to better describe the exact wording used by the survey in describing race/ethnicity.

This has been added.

A very important limitation is the inability to differentiate between screening and diagnostic tests. Any information from other studies that describes the extent to which these tests are for screening would be very helpful.

The primary focus of this study was not determination of the “true” rates for screening for blacks and whites, but determination as to whether the black-white gap in receipt of these procedures differs depending on whether they are based on self-report (regardless of whether these were for screening or diagnostic purposes). However, to better distinguish diagnostic from screening procedures, we undertook two additional analyses. First, we restricted the claims analyses to codes used only for screening. Secondly, we excluded subjects who reported the relevant cancer from the analyses. Both analyses yielded very similar results to those presented.
Finally, most analyses of preventive services conducted by health services researchers using data from the National Health Interview Survey (NHIS) or the Behavioral Risk Factor Surveillance System (BRFSS) do not clearly distinguish between preventive and diagnostic screenings. These include analyses conducted by researchers from CDC’s Center for Chronic Disease Prevention and Health Promotion and also the National Cancer Institute. Similarly, recent analyses that compared rates of mammography from Medicare claims data to self-report data from NHIS and BRFSS did not make this distinction. As we discuss below, it very challenging to reliably do so because of the tendency of providers to justify tests ordered for screening purposes using diagnostic codes.

In one of the above studies, the authors wrote: "To monitor progress in screening for this study, we included tests that were conducted for any purpose, not just as part of a routine examination. We did this for several reasons. First, a person who had the test within the recommended interval has been screened appropriately and ordinarily does not need another test until the recommended time interval has passed. In addition, the baseline and progress toward Healthy People 2010 national cancer screening objectives are measured in this manner."  

The Statistical Analyses section should describe the use of kappa.  
This has been added.

Table 3 should describe what factors are being adjusted for in a footnote. Additionally, it should be clarified why the sample sizes are different in the unadjusted and adjusted analyses.

The covariates included in the adjusted model for Table 3 are listed in a footnote below the Table. The sample sizes for the adjusted analyses are smaller due to missing data for various covariates included in the models.

REVIEWER 2
Major Compulsory Revisions

1. The authors do not provide a rationale for their study objectives to examine racial/ethnic disparities of preventive procedures using both self-report and Medicare claims or a conceptual basis for their hypothesis. As a result, the second paragraph in the Background section comes as surprise to the reader. This section needs to be revised, including a review of the relevant literature in a somewhat more elaborate fashion; the identification of gaps in knowledge; the contribution of the present study to the literature; as a conceptual basis for the study objectives and hypotheses.

This has been added.

2. The extent to which the measures derived truly reflect preventive/screening services is not clear. The authors state “Because of challenges in distinguishing screening from diagnostic procedures, both screening and diagnostic codes were included.” The methodological challenges are indeed well documented in the literature; however numerous studies have attempted to make this distinction, including studies that they have cited (e.g. Cooper and Koroukian). When no distinction can be made between
screening and diagnostic procedures then one wonders how the authors justify keeping the focus in this study on preventive/screening services. The authors either need to incorporate more elaborate algorithms attempting to make such distinctions or drop the term preventive from the study altogether.

The reviewer raises a critical issue here that generated much discussion among the authors; please excuse this lengthy response. First, our primary interest is on focusing on race/ethnicity based *discrepancies* between self-report and claims data. Our starting point is that self-report data (which is what is currently used to monitor *disparities*) have suggested a decrease (even elimination) in *disparities*. It is important to know the extent to which “screening” is distinguished from “diagnostic/treatment” testing in national self-report data. Several of the national surveys (including the MCBS we used, the BRFSS, and the Medical Expenditure Panel Survey) do not attempt to identify non-screening prevention procedures. A follow-up question in NHIS asks for respondents to indicate which of several reasons the procedure was done for, this exclusion has not been applied by researchers in the studies cited above nor has it been applied in the National Center for Health Statistics publication, *Health, United States, 2005*. We suspect that one reason for this it that it is very difficult to reliably distinguish screening from diagnostic testing in self-report surveys. Thus, most of the self-report literature implies or states explicitly that the testing is for “screening” —when a small but significant proportion of the testing is not in fact for screening. There are important implications of this problem that we now highlight in the paper. In the reviewer’s own work it is noted that some of the disparity observed for screening (using an algorithm based on claims data) disappears when diagnostic/treatment testing is considered. The authors astutely observed that this is consistent with minorities presenting later for cancer treatment. This has profound implications for interpreting the meaning of race/ethnicity disparities observed (or not) in the self-report literature. Since that literature does not distinguish between screening and non-screening, and a higher proportion of testing in minorities is likely for non-screening purposes, it is likely that the self-report literature underestimates the actual “screening” disparities. We appreciate the reviewer’s comments as it has suggested to us the need to more fully address this problem—which we do in the paper. However, because our primary aim is to examine differences in disparity estimates based on self-report data and claims data, we prefer to retain the originally reported analyses that compare self-reported procedures with those determined from claims data without exclusions for diagnostic testing. In addition, it is worth noting that there is no entirely satisfactory claims-based algorithm (both high sensitivity and specificity) for determining whether a procedure was for screening. For example, in our health system, identifying a symptom (such as bloating or constipation) is the only way to bypass the logjam delay for getting a “screening” colonoscopy. Similarly, Medicare only reimburses for Pap smear screening biennially. Physicians who wish to screen their Medicare patients more often will often submit claims for these procedures using diagnostic codes resulting in some misclassification error and lowering the estimated rates of screening.
As noted above, we did conduct some analyses restricted to those not reporting cancer and found similar results. Our preference is to report the full analyses, since these parallel most closely the self-report analyses presented elsewhere. However, at the editor’s discretion, we could present these more limited analyses.

The reviewer suggests the alternative of dropping the term “preventive” altogether. We think this may not be an optimal solution to the problem for two main reasons. First, our goal is to be consistent with most of the self-report literature that uses the term preventive, even though, as noted, it is not clear that the procedures are done for exclusively for screening purposes. Second, there is no convenient alternative term to link these tests together. We now note this dilemma in the introduction to alert the reader to the problem. We have been more careful throughout the manuscript to use the phrase “preventive procedures” - to imply that we are referring to a procedure usually done for preventive purposes and have removed the phrase “preventive care.” We would be happy, at the editor’s discretion, to revisit this issue and use a different term.

3. The claims based rates presented in figure 1 seem to be very high for some procedures. For example, the authors report a claims-based rate of colorectal cancer screening of 42.5% among whites and 37.5% among minorities. Based on other published studies, one would expect to see one third of the reported rates (see, for example, KO C et al and report by the General Accounting Office). This raises the question of whether claims were unduplicated across different claims files (Outpatient Standard Analytic File and Provider Supplier File) at the individual level by date of service. One also wonders if Medicare beneficiaries participating in the MCBS database would be more sensitized to the use of preventive services by taking part in the survey and hence higher rates observed in this study. And if that is indeed the case, then the question is whether these results can be generalized to the Medicare population at large. The authors need to compare and contrast their findings with that of comparable studies and the results of this exercise need to be included in the Discussion section of the manuscript.

We appreciate this comment. Indeed, our analytic algorithm looked for duplication within 30-day intervals, so that if, among any of our claims’ sources, we found a billing record for the same procedure in any 30-day period, the claim was only counted once. However, upon re-analysis, we derived rates of 30.1% for whites and 20.4% for blacks for any type of CRC screening. More specifically, we derived rates 16.4% for whites and 9.5% for minorities for FOBT and 13.7% and 10.9 for combined colonoscopy/sigmoidoscopy. These rates are slightly lower than overall (diagnostic, surveillance, and screening) rates reported by Cooper and Koroukian of 18.2% and 11.9% for FOBT and 10.4 and 8.9% for combined colonoscopy/sigmoidoscopy for whites and minorities respectively. Despite our use of liberal criteria for screening, our derived rates were appreciably smaller than self-reported rates from the MCBS.

Minor Essential Revisions

1. Justify the cutpoint for $25,000 in categorizing annual income.
This is the cut-off used in original MCBS. Respondents were also queried using a series of income categories in $5,000 increments up to >$50,000 year, but response rates were appreciably lower than those to question using a $25,000 cut-off. Substitution of these ordinal categories for the binary ones did not alter our results.

2. Under limitations, the authors raise the issue of limited statistical power in some of the analyses. While this is apparent in the tables, the authors should list the analyses in which power may have been compromised, and further include cautionary notes relative to some of the findings.

This has been added.


Medicare began paying for screening colonoscopy in July, 2001. Medicare began paying for screening with flexible sigmoidoscopy for average-risk persons and colonoscopy for high-risk persons in January 1998. Thus, prior to July 2001, providers could not bill Medicare for screening colonoscopy within the general population; all such bills had to either indicate the individual was high risk or testing was being done for diagnostic purposes only.

Sincerely,

Kevin Fiscella, MD, MPH
REFERENCES


