Summary:
The authors present an analysis of a demographic health survey paired with home GPS coordinates in South Africa (National Income Dynamics Study) to assess relationships between distance from health centers, race, socioeconomic status, and health care access. Primary analyses include 1) general description of how sociodemographic characteristics differ by relative distance to health facilities; and 2) correlates of health care access (recent visitation and use of nearest health facility) among a sub-sample of Black Africans in low income strata. While well written, of sound message and a critical topic for program planning and health service research, their manuscript has some limitations, which can likely be mitigated with further analysis considerations.

Major Comments:
1) Why did the authors limit their estimation sample to Black African respondents in the three lowest income strata? It seems the goal of this report is to the hypothesis that race and/or income are associated with distance to health centers and have significant impacts on health care access. Thus, a more valid analysis might include all races and income strata, and to estimate how each of these characteristics is correlated with their outcomes of interest. If the sample is restricted to this sub-group, it is inherently impossible to measure the effect of race or broad income inequality (highest vs lowest strata) on health care access.

2) Many of the conclusions cannot be derived from the restricted analyses presented. For example, the authors state that “respondents in the poorest income quintiles are more likely to live further from the nearest health facility,” – yet aside from a Figure, this data is not shown (and not tested statistically); “we find distance decay in health consultations and attended births, which is largest for the poorest Black African respondents,” – yet again there are no presented data on attended birth use by income or distance to clinic. Additionally all statements about income, race and access to care have only been made within a sub-population of Black Africans in the lowest income strata. Statements comparing health access and distance to clinic by race and income should be supported by models that include all races and show comparative differences by race and income.

3) Overall – while informative, I would suggest a more expansive group of analyses. The authors select rational health indicators with univariable (by race) comparisons presented in Table 1. But they restrict their multivariable analyses
to black race, selected income quintiles only, and select only two health indicators for Table 3 (having a health consultation in past year) and Table 4 (using the closest health facility). Importantly, neither of their multivariable model outcomes of interest are measures of distance to clinic, so definitive predictors of distance to clinic are not made. Moreover, because they do not include race or the full range of income in their multivariable analyses, the impact of race and income cannot be fully extrapolated. With many thousands of survey respondents, their data is rich and can be exploited to make more specific conclusions. I would suggest the following:

a. Select primary outcomes of interest. I would divide these into 1) distance to clinic as an outcome; and 2) health care utilization (clinic visit in past year, attended birth in past year, etc) and include distance as an explanatory variable in the latter set of analyses

b. Perform both univariable and multivariable analyses (adjusted for explanatory variables that the authors include, such as age, self-reported health, medical insurance, education, household size, income, etc.). While stratification by gender and urban/rural seems reasonable, other variables (e.g. income, race, self-reported health status) should be included in all analyses to allow for better discernment if the health disparities in race are independent of socioeconomic status, and vice-versa. If such an analyses is limited by colinearity – than these statement should be made. This analysis strategy will enable a clearer understanding if the inequalities in access are income or race related (or both)

c. Assess for by-race interaction terms for the impact of distance to clinic on health outcomes

4) Is HIV data available (self-reported if not seroprevalence) in their datasets? Given the adult prevalence of HIV and its widespread impact on health care utilization, dependence, etc. it would seem a likely confounder for much of their analysis. This is particularly important for their Table 4 analysis for correlates of use of the nearest health facility.

Minor Comments:

1) The data, which derive from 2008, while informative, are somewhat out of date. A mention of this as a potential limitation in the discussion section seems warranted. These relationships might have changed in numerous ways since the time of the data collection. Most importantly, rapid scale-up of ART provision along with recent national prioritization to decentralize HIV care from PEPFAR to government clinics is likely to determine health care utilization in important ways.

2) Figure 1 is difficult to interpret. Please label both the x and y-axis. Also it appears the y-axes are on different scales, which is misleading

3) Table 1 proportions can be compared statistically (using ANOVA or chi-squared tests) to estimate differences in target characteristics (distance to clinic, self-reported health, etc)

4) The columns in Table 3 and 4 are not labeled and difficult to interpret

5) Figures 3 and 4 should likely have y-axis with an origin at 0%
6) The conclusion statement about quality of health care, while likely accurate, cannot be supported by any available data in this manuscript.

7) The manuscript would likely benefit from more discussion of potential reasons for (perceptions of quality of care, see Ware et al, Plos Medicine), stigmatization of disease states and its impact on health seeking behavior (see Tsai et al, Plos Medicine), and considerations of health service availability by clinical site (e.g. HIV only provided at certain centers)

8) The manuscript would likely benefit from further considerations of strategies to improve distance related health disparities. When health care sector areas would need to be strengthened (human resources, decentralization of services, greater access to specialist, transportation reimbursements, etc)

9) How was income assessed? Were wealth strata weighted to national standards are divided within the cohort?

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests