Author's response to reviews

Title: Rating Neighborhoods for Older Adult Health: Results from the African American Health Study

Authors:

Elena M Andresen (andresen@phhp.ufl.edu)
Theodore K Malmstrom (malmsttk@slu.edu)
Fredric D Wolinsky (fredric-wolinsky@yiowa.edu)
Mario Schootman (mschootman@im.wustl.edu)
J Philip Miller (jphiliomiller@wustl.edu)
Douglas K Miller (dokmille@iupui.edu)

Version: 3 Date: 9 August 2007

Author's response to reviews: see over
We are pleased to submit our revised manuscript “Rating Neighborhoods for Older Adult Health: Results from the African American Health Study.”

We revised and edited the manuscript following the second reviews. Specific responses for both reviewers are listed below. We appreciate the suggestions and found them to be useful. There is a modest increase in word length to address the issues reviewers asked us about, including more detail about our earlier paper on different observer rating scale. We now provide more detail in both the Background and Discussion sections.

One of our new analyses raised a new concern: this is the analysis we conducted on the potential for interaction effects between the two neighborhoods and interviewer/raters (n=18). We conducted this test to answer questions by the three original reviewers: we are in complete agreement that the results of the number of interaction tests must be viewed with caution, and we felt that it would be inappropriate to draw a conclusion about interaction patterns from these tests. We now enhance this important caveat in the discussion. In addition, we tried two methods of this interaction model with the same results, which we describe for the reviewer; however we retained the method we believe is the most appropriate in the paper.

Sincerely,

Elena M. Andresen, PhD.
Professor & Chair, Department of Epidemiology & Biostatistics
College of Public Health and Health Professions
University of Florida Health Sciences Center
PO Box 100231
Gainesville, FL 32610
352-273-5359 office; 352-273-5365 fax; andresen@phhp.ufl.edu

Reviewer: 1

Minor Essential Revisions
1. The Phases are much clearer in the analysis and results. I recommend adding a brief description of them to the Intro to help the reader understand the exact purpose of this paper.
   We agree and added a brief overview of phases at the end of the Background.

2. The first full sentence on page 13 is unclear (“There were some reversed differences…”).
   This sentence was revised and is much clearer (now mid-page 13).
3. Likewise, the first sentence in the second paragraph on page 13 could use clarification. This could probably use a couple more sentences explaining Table 2 because its results are not as self-explanatory as the other tables. The Results description for Table 2 has been augmented. We also changed the footnote to the table and the title. In the title we now use the terms “Interviewer observed” and “Resident” to clarify that observer ratings are compared to ratings of people who lived in the neighborhood. “Respondent” may be an accurate research term, but since the interviewer observer raters also are a type of respondent, this does obscure the comparison and its results. We appreciate this reviewer’s attention to clarity of language.

4. In the Discussion, perhaps you could further emphasize that the nearly 3 point difference in the scale across suburb/urban neighborhoods was due to “real” differences, not an artifact of measurement. I think that is lost in your discussion of findings and it’s an important scientific point. We agree and have emphasized this by both the placement of this point and with added emphasis.

5. The last full sentence on page 16 could use clarification (“In the only other study…”). We revised this sentence to describe the reference in more detail – this was a study of the same general geographic area (St. Louis) but directed toward rating aspects of environments that might improve physical activity.

Discretionary Revisions

6. Page 12. The factor loading for traffic seems low (probably due to higher variability across neighborhoods). Was the alpha higher when traffic was not included? Dropping any one item reduced, rather than improved the alpha; but very slightly. Dropping the traffic item changed the single-block scale alpha from 0.75 to 0.74. We added this to the results text (page 12).

7. Substantively, I tend to disagree with your statement that the neighborhood effects are not confounded by race particularly for the comparison of interviewer-rated and subject-rated neighborhood status; there may be an unmeasured race effect. Your study assumes that white respondents on the same street would rate their neighborhood the same way. You can’t answer that question with your study design, but it still may exist. We agree that we should limit this statement to the cohort under discussion. Confounding may still exist, and the issue deserves follow-up in multi-race studies.

Reviewer: 2

Major Compulsory Revisions

Introduction

1. Page 5, middle paragraph, line 8, the authors’ state that “this assessment tool had acceptable psychometric properties, but we were not satisfied with its discriminative ability or its inter-rater reliability.” The authors should elaborate on the implications of the discriminative ability or the
inter-rated reliability of the scale for studies using the scale. The same will apply to the next sentence state that ‘Further, we found a marked tendency for raters to choose “good” ratings for most items.’ We added details about earlier this study that will help explicitly set the rationale for the paper under review. The details include a description of the rating problems of the earlier scale and the large proportion of ratings as “good” (scored as 2 points), the impact on the scale distribution, and also details about the interviewer effects in the earlier scale.

2. The aim of the study could be more specific regarding what exactly the authors sought in this study. Both the previous study and the goals of the paper under review were changed in the Background as both reviewers suggested.

3. Methods. The authors should consider being more specific on explaining the purpose or outcome expected for each phase and how the analytical phases are connected. This clarification will make the MS’ results easy to follow and can improve the MS. We agree, and these issues are part of the new Background. We also revised the methods section on the phases.

4. Results. For phase 3, the authors stated that they tested 17 interactions (end of page 14). The validity of a model with 17 interactions plus 18 or so main effects is highly questionable. Thus, the interaction findings need to be interpreted with caution. We agree with this caution. We preliminarily constructed a single model with an interaction term for interviewers and catchment area. This reduced the number of tested terms, but was not easy to interpret for the direction of interaction of interviewer by catchment area. The tests we conducted that are described in the manuscript produced the same two interviewers as “different” and we also found that there was not a consistent direction in their “interaction.” We have added language about caution in the discussion.

5. Discussion. In the Introduction, the authors state that ‘this assessment tool had acceptable psychometric properties, but we were not satisfied with its discriminative ability or its inter-rater reliability.’ Thus, what do the authors learn from this analysis? What can this study contribute to this field? Were there any limitations on the analyses? Details of the simple, earlier five-item rating scale are now provided and contribute more clearly to the rationale for the present manuscript, and to the Discussion. We edited the list of limitations and strengths based on a number of points raised by both reviewers, above.