Professor Stewart,
You wrote:
"There were several counties incorrectly specified as southern in the coding scheme we employed."
Thank you for that information.
You wrote:
"Moreover, if possible, would it be possible to schedule a conference call? I was responsible for the data analysis. I would be happy to discuss the concerns you raised."
I accept your offer to discuss my concerns. However, I do not believe a telephone call is the appropriate mode. As a scientific discussion, this is best conducted in writing and with your coauthors included. (I returned them to the email chain.) There seem to be new irregularities in the corrected (published) version of Mears et al. (2019). Can you please explain them? It's unclear how removing respondents accidently coded as Southerners in the initial article (N = 1,441 to N = 1,301, a difference of 140) would alter the data in the following ways:

Incongruent Changes in Mears et al. (2019)

There are four changes to descriptive statistics (online-first vs. published sample) that seem mathematically and logically improbable. This web link is to a picture of the two tables, with discrepancies highlighted in yellow. Underneath the web link, I describe each discrepancy:

https://imgur.com/a/XYauxUK

The average age of the sample increased by almost 8 years (from 40.03 to 47.85). Mathematically, this does not seem possible. My calculations suggest that even if all 140 dropped respondents were 18 years old, which is the youngest age in the sample, the average age would increase only by two years, to 42.40. In fact, even if all 140 dropped respondents were five years old, the average age would increase only to 43.80, four years less than the published mean of 47.85. Here is how I am calculating this: multiply the initial sample's size (1,441) by its mean age (40.03) to get the sum of all ages in that sample (57,683.23). Multiple the number of respondents dropped, here

140, by a chosen age, say 18 years old, to get the sum of ages in the dropped subsample (2,520), if the dropped respondents were that age. Subtract that number from the total sum of ages in the initial sample to get the sum of ages in the reduced sample (57,683.23 - 2,520 = 55,163.23). Divide that number by the new sample size, here 1,301, to get the mean age in the reduced sample (55,163.23/1,301 = 42.40).

In the online-first sample, mean concentrated disadvantage is 3.59, and the mean homicide rate is 9.26. In the published sample, concentrated disadvantage is over half a standard deviation higher (4.71), but homicide is over half a standard deviation lower (6.12). So, the counties in the published sample are much more disadvantaged, on average, but also have much lower homicide rates. This is unlikely, because concentrated disadvantage is strongly positively related to homicide (Sampson, Raudenbush, & Earls, 1997).

The increase in concentrated disadvantage is surprising for another reason. Despite the published sample living in more disadvantaged areas, on average, and despite it having 140 fewer respondents, its family income is identical to the online-first sample's in both mean and standard deviation (Mean = $55,314, s.d. = $10,347). Furthermore, both education and employment should affect income, but the online-first and published samples have identical incomes, despite having different proportions of college educated (44% vs. 48%) and employed (49% vs. 52%) respondents. These differences alone should result in differences in mean family income.

Dropping respondents from a sample should never widen the min/max range of values (if anything, it should narrow it), because all values remaining after a data drop

were there before the drop. A total of 140 respondents were dropped from the online-first sample to get the published sample. After dropping these respondents, the maximum value of population structure increased to 15.61 (from 9.41), the maximum value of concentrated disadvantage increased to 11 (from 9), and the minimum value for “black growth” fell to -.07 (from -.06). Mathematically, none of these changes seem possible. Because the smaller published sample is nested within the larger online-first sample, the min/max range in the published sample should fall within the min/max range of the online-first sample.

Respectfully, John Smith

On Thursday, May 23, 2019, 10:17:12 PM EDT, Eric Stewart <redacted> wrote:

Mr. Smith,

Apologies for the delayed response. A new semester of classes just started. There were several counties incorrectly specified as southern in the coding scheme we employed. In effect, these counties should have been specified as non-southern locations. Once we identified the coding errors, we re-estimated all models.

Moreover, if possible, would it be possible to schedule a conference call? I was responsible for the data analysis. I would be happy to discuss the concerns you raised.

Eric

On May 19, 2019, at 11:17 AM, John Smith <redacted> wrote:

Professor Stewart,
Thank you for agreeing to look into the irregularities. Because you corrected Mears et al. (2019) already, will you please tell me what coding errors necessitated the post-online-publication changes to that article?
Respectfully,
John Smith

On Saturday, May 18, 2019, 05:53:44 PM EDT, Eric Stewart <redacted> wrote:

Hello Mr. Smith:

Thank you for bringing these concerns to our attention. We will investigate them. And if there in fact are problems, we will address them.

Eric Stewart