A Reproduction of the Results of Onyike et al. (2003)

Nicholas J. L. Brown
University of Groningen

Jan van Rongen
Independent consultant

Jakob van de Velde
Ghent University

Matt Williams
Massey University

Abstract

Onyike et al. (2003) analyzed data from a large-scale US-American data set, the Third National Health and Nutrition Examination Survey (NHANES-III), and reported an association between obesity and major depression, especially among people with severe obesity. Here, we report the results of a detailed replication of Onyike et al.'s analyses. While we were able to reproduce the majority of these authors' descriptive statistics, this took a substantial amount of time and effort, and we found several minor errors in the univariate descriptive statistics reported in their Tables 1 and 2. We were able to reproduce most of Onyike et al.'s bivariate findings regarding the relationship between obesity and depression (Tables 3 and 4), albeit with some small discrepancies (e.g., with respect to the magnitudes of standard errors). On the other hand, we were unable to reproduce Table 5, containing Onyike et al.'s findings with respect to the relationship between obesity and depression when controlling for plausible confounding variables—arguably the paper's most important results—because some of the included predictor variables appear to be either unavailable, or not coded in the way reported by Onyike et al., in the public NHANES-III data sets. We discuss the implications of our findings for the transparency of reporting and the reproducibility of published results.

Keywords: Body mass index, Body Weight, Depression, Obesity, Weighted surveys

Background

In the spring of 2016, the first author (Nick Brown) had a plan, as part of his PhD studies, to perform some analyses using the U.S. American Third National Health and Nutrition Examination Survey (NHANES-III) data set. To improve his understanding of the NHANES data and code books for this project, Nick decided to download an article that was based on the same data set and attempt to reproduce the results. Somewhat arbitrarily, he chose the widely-cited article on the topic of the relation between obesity and depression by Onyike et al. (2003).

Onyike et al. (2003) reported two major findings. First, obesity—defined as a body mass index (BMI) above 30—was associated with past-month depression in women (but not men). Second, severe obesity—defined as a BMI above 40—was associated with past-month depression in men and women combined. At the time, Nick found that he could not reproduce the results from this article, other than some of the most basic descriptives. He wrote to the lead/corresponding author of the Onyike et al. article, Dr Chaidi Onyike, but after a brief exchange the correspondence ceased, and he decided to put the exercise of reproducing Onyike et

al.'s results to one side for the time being.[1]

In March 2018 Nick resurrected the project with a blog post (Brown, 2018) asking for volunteers to help with independent reanalyses. Several people responded, and three stayed on board long enough to contribute substantial amounts of code and insights. This article, of which those three people are the second through fourth authors, is the result of that exercise.

At an early stage of our reanalyses, it became clear the one of the main reasons why Nick had not been able to reproduce Onyike et al.'s (2003) results was his failure to notice that the survey results were weighted to make the sample as representative as possible of the U.S. population. However, once that elementary problem had been overcome, several other issues emerged with the design choices of Onyike et al.'s study as well as each of the individual tables of results that could not be so easily explained. We discuss these issues in the following sections. Further details, in particular concerning the different ways available to calculate standard errors, can be found in a technical note by Jan van Rongen, which we have made available, along with our analysis code, at https://osf.io/j32yw.

In this article, our primary focus is on reproducing the results reported by Onyike et al. (2003), rather than evaluating the appropriateness of their analyses or the validity of the study as a whole. Nevertheless, we have included some brief commentary at a few points where the analyses conducted by Onyike et al. seem to have had clear problems.

Data processing

The NHANES-III survey upon which Onyike et al.'s (2003) paper is based was conducted in the United States in two phases between 1988 and 1994. Data sets and documentation for NHANES-III are openly accessible from this link. It appears that Onyike et al. used data from three of the four main data sets produced by the survey: "Household Adult", "Household Youth", and "Examination". Our initial data processing steps involved downloading and merging these data sets, applying the inclusion and exclusion criteria specified in Onyike et al. (see p. 1141), and creating derived variables (e.g., age and BMI categories). Most of this process was relatively straightforward and we do not describe the steps in detail here; see the analysis script in our OSF project for further information. However, it is worth describing how we created the principal outcome variables (prevalence of depression at various time points). Onyike et al. classified participants as having had a diagnosis of depression (a) at any point in their lifetime, (b) in the past month, (c) in the past year, and (d) recurrently. The NHANES-III examination data contains several depression-related variables. We took a value of 3 for the variable MQPDEP (any lifetime diagnosis of major depressive disorder, except following a bereavement) to indicate "lifetime depression." We assumed that past-month and past-year depression were defined by the variable MQPLDDP, which measures how long ago the last episode was diagnosed, with the two values 51 ("within the last 2 weeks") and 52 ("between 2 weeks and 1 month ago") indicating past-month depression, and these two values together with 53 ("1 to 6 months ago") and 54 ("6 months to 1 year ago") indicating past-year depression. Recurrent depression was taken from the variable MQPDEPRT, with values of 2 or 3 indicating two or more lifetime diagnoses of major depression with any level of severity (NHANES-III, 1996b). The remainder of our manuscript is organized according to the five tables reported in Onyike et al. (2003). While a small quantity of additional statistical information was provided in their main text (which we have not discussed below), these tables appear to contain all of the most important results of their study.

Table 1

We were able to reproduce Onyike et al.'s (2003) Table 1 exactly, apart from the column labeled "β". Onyike et al. did not provide a legend for this column, but we assume, from their remark that "There was sufficient statistical power to test the study hypotheses" (p. 1141), that it is a form of post hoc power calculation (in which case the column label should arguably have been "$1-\beta$"). Numerous authors (e.g., Lakens, 2014) have pointed out that post hoc power calculations amount to little more than a transformation of the p value, so the utility of this column is perhaps questionable. We were unable to reproduce the reported "β" figures, because we do not know how Onyike et al. dealt with the issues of (a) sample weighting and (b) the design effect (i.e., the correlations among clustered observations due to the non-random survey design; cf. NHANES-III, 1996c, pp. 25–27) in their analyses. We refer interested readers to Jan van Rongen's technical note in our OSF repository for more details on this.

Table 2

Table 2 in Onyike et al. (2003) contains demographic information (stratified by gender), with standard errors for means and percentages falling in various groups (e.g., age and ethnicity).

---

[1]On July 11, 2020, we wrote again to Dr Onyike, enclosing a copy of the preprint version of the present article, and inviting him to comment. However, as of August 28, 2021, we have received no reply of any kind.

Table 1

Reproduction of Onyike et al.'s (2003) Table 1.

Hypothesis and sample	n1	n2	p1	p2	n2/n1	β
Hypothesis A: Obesity (body mass index ≥30) is associated with depression.						
All respondents	4,154	1,658	0.028	0.051	0.40	<u>0.985</u>
Females	2,180	1,084	0.038	0.067	0.50	<u>0.943</u>
Males	1,974	574	0.017	0.029	0.29	<u>0.449</u>
Hypothesis B: Class 3 (severe) obesity (body mass index ≥40) is associated with depression.						
All respondents	4,154	267	0.028	0.125	0.06	<u>1.000</u>
Females	2,180	202	0.038	0.130	0.09	<u>0.995</u>
Males	1,974	65	0.017	0.115	0.03	<u>0.947</u>

Note. Underscored values are different from those reported in Onyike et al.'s (2003) Table 1. See Onyike et al.

(2003, p. 1142) for column label legends.

Somewhat confusingly, Onyike et al.'s (2003) Table 2 indicates a sample size of 8,773 (4,745 female, 4,028 male), whereas all of their other results were based on a sample size of 8,410. The difference between these two sets of participants is explained by the fact that in the original sample of 8,773 participants aged 15–39 years who underwent the medical examination and interview (thus meeting the inclusion criteria), the data (height and/or weight) needed to calculate body mass index (BMI) were missing for 25 people. Similarly, sufficient interview data to make a diagnosis of (non)depression were missing for 14 people, and both of these elements were missing for 324 people. Hence, Onyike et al. excluded a total of (25 + 14 + 324) = 363 people from the final sample (8,773 – 363 = 8,410). It is not clear to us why the larger data set was used for describing demographic characteristics, given that it was apparently not used for the rest of the analyses in Onyike et al.'s article.

Although Table 2 is described by Onyike et al. (2003) in their text as displaying "The demographic characteristics of the respondents" (p. 1141), our attempts to reproduce this table suggest that this is not the case: The estimates provided in the table appear to have been produced with sample weighting applied, meaning that they are actually estimates of the demographic characteristics of the U.S. population (and, consequently, substantially *different* from the demographic characteristics of the respondents)[2]. This also means that Table 2 con-

tains a discrepancy between the Ns reported at the top of each column (which show that the sample was 54.1% female) and the percentages reported in the "Gender" breakdown (50.7% female), which refer to the population.

In applying weighting in these and all subsequent analyses we assume that Onyike et al. (2003) used the sample weighting variable WTPFEX6 ("examined sample final weight"). This is not the only weighting variable available in the NHANES-III datasets, but the use of other available weighting options (e.g., WTPFQX6, "interviewed sample final results") gives results that do not match Onyike et al.'s Table 2.

Having identified the weighting variable applied by Onyike et al. (2003), three of us, working independently, found three different ways of calculating the reported standard errors. We established that the JKn jackknife method in the R function *as.svrepdesign* from the *survey* package (Lumley, 2019) produces almost exactly the same standard errors as those reported by Onyike et al. We tentatively assume that the version of Stata used by Onyike et al. produced standard errors for descriptives in the same way. Any remaining minor differences between our values and those of Onyike et al. might stem from the choice of underlying replicate

---

[2]The caption of the table, in contrast, refers to the "characteristics of the study population", so is more consistent with the statistics provided.

designs (see the accompanying Technical Note).

In addition to the general issues outlined above, we had some specific difficulty in reproducing Onyike et al.'s (2003) percentages for ethnicity by gender. After some experimentation, we established that their exact numbers could be reproduced if we used the $N = 8{,}773$ data set for females, and the $N = 8{,}410$ data set for males (in which there were 3,849 male participants). In other words, our earlier note about the use of data prior to exclusions for this table does not apply to the specific case of the ethnicity of males.

A final issue in Table 2 relates to the Education frequencies. Our results for the percentages of participants with more than 12 years of education are slightly lower than those reported by Onyike et al. (2003). On closer examination of the data set, it appears that Onyike et al. counted any numerical value above 12 in the NHANES-III variable HFA8R as representing more than 12 years of education. However, as the NHANES-III Adult data file code book makes clear (NHANES-III, 1996a, p. 96), some participants have the value 88 ("Blank but applicable"—which, confusingly, appears to mean "impossible value", cf. NHANES-III, 1996, p. 21) or 98 ("Don't know") for this variable, which corresponds to missing data. Hence, we believe that our numbers are the correct ones here. Our accompanying code also calculates the education results treating these "missing" values as indicating more than 12 years of education; the results in that case correspond exactly to those reported by Onyike et al.

In sum, while we were able to reproduce most of the numbers reported in Table 2, doing so involved applying analysis decisions that seem at odds with the written description in Onyike et al. (2003), and at least two errors appear to have been made in producing the table. On the other hand, this Table primarily contains descriptive information, so it is not critical to the conclusions of the study.

Table 3

Onyike et al.'s (2003) Table 3 displays how the estimated prevalence of past-month depression varies across BMI categories (with stratification by gender). It provides only point estimates, and no indicators of uncertainty (e.g., confidence intervals, p values). We were able to reproduce this table in its entirety, with the exception of a discrepancy of 0.01 in one percentage (possibly due to different rounding between Stata and R), and an apparent transcription error in the number of people in Obesity class 1, where Onyike et al. reported 910 instead of 981 (a value that Onyike et al. themselves reported correctly in their Table 4).

We disagree, however, with Onyike et al.'s (2003)

choice of label for the third column of this table, "All respondents" (and, by implication, the fourth and fifth columns also, assuming that "Females" and "Males" implicitly carry over the term "respondents" for each sex). The word "respondents" suggests that, looking for example at the third column of the first line of Table 3, 2.79% of *people who responded to the survey* had normal body weight and met the criteria for a past-month diagnosis of depression, whereas in fact this figure of 2.79% represents the *estimate* for the total population of the US based upon the weights and the survey design.

Table 4

Table 4 in Onyike et al. (2003) displays differences in the estimated odds of past-month depression across BMI categories. These differences are expressed in the form of odds ratios comparing the odds of depression in various BMI categories to the odds of depression in participants of normal weight.

We exactly reproduced almost all of the point estimates of the odds ratios in Table 4, with three exceptions: All respondents, Past-year, Obese, where we obtained a result of 1.42, versus 1.41 in Onyike et al.'s (2003) article; Females, Past-month, Obesity class 1 (1.32 versus 1.28); and Females, Past-month, Obesity class 2 (1.84 versus 1.75). The first of these discrepancies might be due to rounding, but it is not clear what could have caused the other two. Overall, however, the level of agreement between our table and the original gives us confidence that our derivation of the four depression category variables from the NHANES-III measures was faithful to that of Onyike et al.

The majority of the confidence interval boundaries in our reproduction of Table 4 were also close to those of Onyike et al. (2003) within a margin of 0.01 or 0.02, suggesting that the method that we chose for determining the standard errors of the odds ratios among the numerous options that the survey package makes available, namely JKn (jackknife for stratified designs; Lumley, 2019), was the one that most closely matches that applied by Onyike et al. However, for the three lines of Table 4 with the smallest sample sizes—the BMI classes "Underweight," "Obesity class 2," and "Obesity class 3" for males, with sample sizes of 99, 125, and 65, respectively—our CIs were even wider than those of Onyike et al., in some cases by a considerable margin. The total number of male participants in those three BMI classes who reported ever being depressed in their lifetime was 5, 3, and 7 respectively (and these numbers were, naturally, even lower for recurrent, past-year, or past-month depression, with just one male participant in the BMI 35–39.9 category having recurrent or past-month depression).

Table 2

Reproduction of Onyike et al.'s (2003) Table 2.

Characteristic	Females(n=4,745)		Males (n=4,028)	
	%	SE	%	SE
Gender	50.6	0.6	49.4	0.6
Age (years)				
15-19	17.2	1.2	17.8	0.9
20-24	20.3	1.3	19.2	1.0
25-29	19.6	1.2	21.0	1.1
30-34	21.5	1.4	22.3	1.3
35-39	21.4	1.2	19.7	1.2
Race/ethnicity				
White	70.0	1.6	71.2	1.7
Black	14.0	1.0	12.1	0.7
Hispanic	12.0	1.2	12.4	1.1
Other	4.1	0.6	4.3	0.6
Education (years)				
0–8	6.8	0.8	7.4	0.6
9–11	19.7	1.1	22.3	1.2
12	33.6	1.1	31.4	1.2
>12	39.4	1.8	38.6	1.7
Marital status				
Married	52.2	1.4	50.9	1.6
Separated/divorced/widowed	11.4	0.7	4.8	0.5
Never married	36.2	1.5	44.1	1.7
Area of residence				
Urban	49.6	5.0	50.4	4.9
Rural	50.4	5.0	49.6	4.9

Notes. Underscored values are different from those reported in Onyike et al.'s (2003) Table 2.

For "Race/ethnicity", the number of males is 3,849; see discussion in the main text.

Table 3

Reproduction of Onyike et al.'s (2003) Table 3.

Relative body weight	No. of participants	% with DIS/DSM-III depression		
		All respondents	Females	Males
Normal weight (BMI 18.5–24.9)	4,154	2.79	3.82	1.67
Underweight (BMI <18.5)	301	3.24	3.82	1.82
Overweight (BMI 25.0–29.9)	2,297	2.42	4.01	1.37
Obese (BMI ≥30)	1,658	<u>5.13</u>	6.74	2.85
Obesity class 1 (BMI 30–34.9)	<u>981</u>	3.55	4.97	1.88
Obesity class 2 (BMI 35–39.9)	410	4.80	6.79	0.83
Obesity class 3 (BMI ≥40)	267	12.51	13.03	11.54

Note. Underscored values are different from those reported in Onyike et al.'s (2003) Table 3.

Numbers in parentheses represent the standard error of the corresponding percentage estimate.

In most cases, the wider confidence intervals in our reproduction do not affect whether the odds ratios reported in Table 4 are statistically significant at the .05 level, with two exceptions:

- The odds ratio for the relationship between BMI (treated as a continuous variable) and past month depression in females is statistically significant in Onyike et al.'s (2003) Table 4, 95% CI [1.03, 1.06], but not in our reproduction, 95% CI [0.99, 1.04].

- The odds ratio for the comparison of the prevalence of past-month major depression between obesity class 3 and normal weight participants in the male subsample is statistically significant in Onyike et al.'s Table 4, 95% CI [1.03, 57.26], but not in our reproduction, 95% CI [0.12, 486.2]. As mentioned above, there were only three male participants in obesity class 3 for this comparison; it does not seem implausible that minor variations in calculation methods between statistical

software packages could cause substantial differences in their outputs for such small subsamples.

These two discrepancies nevertheless relate to relatively ancillary findings that were not emphasized in Onyike et al.'s (2003) abstract or discussion.

Table 5

We were unable to reproduce Onyike et al.'s (2003) Table 5 because several of the covariates that these authors claimed to have included were either not available in the NHANES-III data set that we downloaded, or were calculated in an unclear way. Specifically:

- We were unable to find any measure of the use of psychiatric medicine in the NHANES-III data set or code books.

- We have no way to determine the criteria used by Onyike et al. to categorize participants' alcohol use as *None*, *Moderate*, and *Abuse*, based on the six variables (MYPF1, MYPF2, MYPF3S, MYPF4,

Table 4

Reproduction of Onyike et al.'s (2003) Table 4.

Population and BMI category	No. of participants	Past-month major depression		Past-year major depression		Lifetime major depression		Recurrent major depression	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
All respondents									
BMI (continuous variable)	8,410	1.05	1.01, 1.09	1.03	0.99, 1.06	1.02	0.99, 1.05	1.01	0.97, 1.05
Normal weight (BMI 18.5–24.9)	4,154	1.00§		1.00§		1.00§		1.00§	
Underweight (BMI <18.5)	301	1.17	0.49, 2.77	1.39	0.66, 2.92	1.35	0.74, 2.45	1.21	0.54, 2.69
Overweight (BMI 25.0–29.9)	2,297	0.86	0.53, 1.40	0.84	0.53, 1.32	0.93	0.65, 1.33	0.84	0.54, 1.28
Obese (BMI ≥30)	1,658	1.88	1.03, 3.43	1.42	0.86, 2.33	1.22	0.82, 1.81	1.13	0.73, 1.76
Obesity class 1 (BMI 30–34.9)	981	1.28	0.65, 2.53	1.01	0.55, 1.84	0.87	0.55, 1.38	0.78	0.47, 1.29
Obesity class 2 (BMI 35–39.9)	410	1.76	0.78, 3.97	1.67	0.92, 3.06	1.39	0.78, 2.46	1.41	0.72, 2.77
Obesity class 3 (BMI ≥40)	267	4.98	2.07, 11.98	2.92	1.28, 6.63	2.60	1.39, 4.86	2.28	0.92, 5.67
Females									
BMI (continuous variable)	4,561	1.05	1.01, 1.08	1.02	0.99, 1.05	1.02	0.99, 1.04	1.00	0.97, 1.03
Normal weight (BMI 18.5–24.9)	2,180	1.00§		1.00§		1.00§		1.00§	
Underweight (BMI <18.5)	202	1.00	0.38, 2.62	1.36	0.61, 3.02	1.20	0.59, 2.43	1.03	0.40, 2.62
Overweight (BMI 25.0–29.9)	1,095	1.05	0.65, 1.72	0.81	0.54, 1.21	0.94	0.66, 1.34	0.71	0.45, 1.12
Obese (BMI ≥30)	1,084	1.82	1.02, 3.25	1.29	0.81, 2.07	1.12	0.78, 1.61	0.97	0.64, 1.49
Obesity class 1 (BMI 30–34.9)	597	1.32	0.61, 2.86	0.90	0.45, 1.80	0.74	0.43, 1.28	0.68	0.37, 1.25
Obesity class 2 (BMI 35–39.9)	285	1.84	0.71, 4.75	1.66	0.79, 3.46	1.41	0.74, 2.70	1.40	0.67, 2.94
Obesity class 3 (BMI ≥40)	202	3.78	1.67, 8.55	2.19	0.97, 4.87	2.15	1.19, 3.87	1.36	0.60, 3.13
Males									
BMI (continuous variable)	3,849	1.06	0.97, 1.16	1.04	0.98, 1.10	1.02	0.97, 1.07	1.03	0.96, 1.10
Normal weight (BMI 18.5–24.9)	1,974	1.00§		1.00§		1.00§		1.00§	
Underweight (BMI <18.5)	99	1.09	0.13, 9.28	0.57	0.07, 4.95	1.06	0.23, 4.90	1.12	0.04, 29.83
Overweight (BMI 25.0–29.9)	1,202	0.82	0.35, 1.94	1.08	0.56, 2.07	1.16	0.65, 2.06	1.25	0.64, 2.45
Obese (BMI ≥30)	574	1.73	0.52, 5.71	1.54	0.71, 3.36	1.28	0.67, 2.47	1.40	0.57, 3.45
Obesity class 1 (BMI 30–34.9)	384	1.13	0.40, 3.17	1.22	0.59, 2.54	1.14	0.61, 2.13	1.00	0.40, 2.47
Obesity class 2 (BMI 35–39.9)	125	0.49	0.00, 3.1e8	0.99	0.16, 5.97	0.66	0.11, 3.94	0.71	0.00, 1.9e9
Obesity class 3 (BMI ≥40)	65	7.68	0.12, 486.2	4.53	0.41, 50.07	3.26	0.40, 26.54	5.15	0.53, 50.23

Note. Underscored values are different from those reported in Onyike et al.'s (2003) Table 4. §: Reference category.

7

MYPF5S, and MYPF6S) that correspond to participants' responses to questions that were about their alcohol consumption in the NHANES-III interview.

- Onyike et al. classified participants as (a) current smokers, (b) former smokers, or (c) those who had never smoked. The NHANES-III survey and examination data sets contain a number of items related to the smoking of cigarettes, cigars, and pipes; it is not clear how these were combined to arrive at Onyike et al.'s three-way classification.

- We do not understand why the five categories (*Excellent*, *Very good*, *Good*, *Fair*, *Poor*) for physician's health rating from the NHANES-III examination (variable PEP13A) were collapsed into just three categories (*Excellent*, *Good*, *Fair/poor*).

- We do not understand why the four race/ethnic categories from Table 2 were collapsed into three in Table 5, with "Hispanic/other" apparently being used as an omnibus category for anyone who was not classed as "White" or "African-American" (this last category apparently being a synonym for "Black" from Table 2).

We could, of course, have reproduced the table with these covariates either omitted or guessed at, but a comparison of the results with the published table would probably not have been very meaningful.

Discussion

Our efforts to reproduce Onyike et al.'s (2003) analyses were made easier by the fact that the underlying data set was openly accessible and extensively documented (the NHANES-III documentation consists of many hundreds of pages for each data set). This is in contrast to the situation facing researchers who wish to reproduce articles for which the data are less thoroughly documented or simply not available for re-analysis at all. Despite this, however, it was difficult for us to reproduce many of Onyike et al.'s tables, because we did not know how all of the choices that these authors made in analyzing the data. The fact that our reanalysis was so challenging even in this seemingly favorable scenario speaks to the importance of sharing not only data and descriptions of analyses, but also the original code (typically in the form of scripts in the language of a statistical software package) that was used to process and analyze the data. It is only with access to this code that readers and reviewers can obtain full insight into how the data were actually analyzed.

A number of positive changes in the process of analyzing scientific data and publishing the results of those analyses have taken place in the 18 years since Onyike et al.'s (2003) article was published. First, the widespread dissemination and adoption of free software such as R (R Core Team, 2018) and its associated packages has made powerful computing tools and associated support resources available at essentially no cost to anyone with access to a quite modest desktop or laptop computer. Second, organizations such as the Open Science Foundation (https://osf.io/) now make it easy for authors to share their analysis code and (depending on licensing arrangements and confidentiality issues) data. Third, helped by the improvements mentioned in the two previous points, the sharing of code and data so that other researchers may reproduce and possibly extend one's results is rapidly becoming a standard part of publishing a scientific article (e.g., Lindsay, 2017). All of those developments have played their part in our replication efforts and the writing of the current article.

In our reproduction, we were able to reproduce most of the figures in Onyike et al.'s (2003) Tables 1 and 2, although the analyses necessary to reproduce Table 2 are somewhat inconsistent with the written description in the article (cf. the issue of the "demographic characteristics of the respondents"), and include what appear to be at least two data processing errors. Nevertheless, Tables 1 and 2 represent primarily descriptive information rather than statistics bearing on Onyike et al.'s research questions. For Tables 3 and 4 (representing bivariate relationships between BMI and various operationalizations of depression), we were able to reproduce the reported statistics, albeit with some minor discrepancies. On the other hand, we were completely unable to reproduce Table 5. This table represents arguably the most crucial statistical output of the study, in that it presents information about the relationship between BMI and depression while controlling for the variables that Onyike et al. considered to be plausible confounds. Our inability to reproduce the statistics in this table does not mean that Onyike et al.'s results are invalid—indeed, they are entirely congruent with the findings of subsequent systematic reviews and meta-analyses, such as those by Luppino et al. (2010) and Pereira-Miranda et al. (2017)—but it does suggest that they were presented without sufficient information to permit direct replication.

Despite the issues we have raised in the present article, we do not believe that Onyike et al.'s (2003) article is severely flawed; certainly we do not think that it is atypical of the research that was being published at the time. Nor do we think that an extensive corrigendum is required, although perhaps a brief note could be added to the published article to correct the most obvious errors that we have identified and add suffi-

cient information about the data preparation and analysis process to allow the reproduction of the reported results. Our take-home message for researchers is, rather, a more general one: Even with a carefully curated data set such as NHANES-III, the process of data analysis requires precision and care, preferably with multiple sets of eyes and the sharing of code (and, where they are not already public, data) to allow for computational reproducibility (Donoho, 2010) of their findings. We believe that the time needed for the reader of an article to reproduce the calculations in a published paper ought to be measurable in minutes, not months.

Author Contact

Corresponding author is Nicholas J. L. Brown. Author contact: nicholasjlbrown@gmail.com

Conflict of Interest and Funding

Author Contributions

All four authors analyzed the data independently. Nicholas J. L. Brown wrote the paper and the other authors provided critical revisions. All authors approved the final version of the manuscript.

Open Science Practices

This article earned the Open Materials badge for making the materials available. This is a commentary that focused on reproducing the findings of a published article, and as such there are no (new) data. It was not pre-registered. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

Brown, N. (2018, March 13). Announcing a crowdsourced reanalysis project [Weblog post]. Retrieved August 28, 2021 from https://steamtraen.blogspot.com/2018/03/announcing-crowdsourced-reanalysis.html

Donoho, D. L. (2010). An invitation to reproducible computational research. *Biostatistics*, *11*(3), 385–388. https://doi.org/10.1093/biostatistics/kxq028

Lakens, D. ((2014, December 19). Observed power, and what to do if your editor asks for post-hoc power analyses [Weblog post]. Retrieved August 28, 2021 from https://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html

Lindsay, D. S. (2017). Sharing data and materials in *Psychological Science*. *Psychological Science*, *28*(6), 699–702. https://doi.org/10.1177/0956797617704015

Lumley, T. (2019). Package 'survey', v. 3.35-1. https://cran.r-project.org/web/packages/survey/survey.pdf

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Penninx, B. W. J. H., & Zitman, F. G. (2010). Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. *Archives of General Psychiatry*, *67*(3), 220–229. https://doi.org/10.1001/archgenpsychiatry.2010.2

NHANES-III. (1996a). *Third National Health and Nutrition Examination Survey (NHANES III), 1988–94: NHANES III household adult data file documentation*. http://www.nber.org/nhanes/ftp.cdc.gov/pub/Health_Statistics/NCHS/nhanes/nhanes3/1A/adult-acc.pdf

NHANES-III. (1996b). *Third National Health and Nutrition Examination Survey (NHANES III), 1988–94: NHANES III examination data file documentation*. http://www.nber.org/nhanes/ftp.cdc.gov/pub/Health_Statistics/NCHS/nhanes/nhanes3/1A/exam-acc.pdf

NHANES-III. (1996c). *Third National Health and Nutrition Examination Survey (NHANES III), 1988–94: Analytic and reporting guidelines*. https://wwwn.cdc.gov/nchs/data/nhanes/analyticguidelines/88-94-analytic-reporting-guidelines.pdf

Onyike, C. U., Crum, R. M., Lee, H. B., Lyketsos, C. G., & Eaton, W. W. (2003). Is obesity associated with major depression? Results from the Third National Health and Nutrition Examination Survey. *American Journal of Epidemiology*, *158*(12), 1139–1147. https://doi.org/10.1093/aje/kwg275

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Pereira-Miranda, E., Costa, P. R. F., Queiroz, V. A. O., Pereira-Santos, M., & Santana, M. L. P. (2017). Overweight and obesity associated with higher depression prevalence in adults: A systematic review and meta-analysis. *Journal of the American College of Nutrition*, *36*(3), 223–233, https://doi.org/10.1080/07315724.2016.1261053

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632