

DISCUSSION PAPER SERIES

IZA DP No. 12575

**Missing Data in Imputed Highest Grade
Completed in the 2015–2018 NBER CPS
Extracts**

Joni Hersch
Fernando Mendoza Lopez
Jennifer Bennett Shinall

AUGUST 2019

DISCUSSION PAPER SERIES

IZA DP No. 12575

Missing Data in Imputed Highest Grade Completed in the 2015–2018 NBER CPS Extracts

Joni Hersch

Vanderbilt Law School and IZA

Fernando Mendoza Lopez

Vanderbilt Law School

Jennifer Bennett Shinall

Vanderbilt Law School

AUGUST 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Missing Data in Imputed Highest Grade Completed in the 2015–2018 NBER CPS Extracts

In 2015, the Current Population Survey (CPS) eliminated three questions related to educational attainment. These questions are used by the NBER to calculate the variable, “Imputed Highest Grade Completed” (*ihigrdc*) in their Monthly Outgoing Rotation Group (MORG) extracts. Imputed Highest Grade Completed provides a convenient measure of years of education and is based on the credential oriented CPS variable that is coded from 31 to 46. Because the NBER continues to use coding that relies on these eliminated questions to calculate *ihigrdc*, this variable has a missing value for 27.5% of the observations in the 2015–2018 extracts. These missing values, in turn, lead to an average Imputed Highest Grade Completed of about 1.2 fewer years after 2014 than would result from using the whole CPS. We informed the NBER of this issue in January 2019; however, because this variable with missing values remains on their website for download as of this writing (July 15, 2019), we are posting this working paper to inform users of these data so that they may address the issue appropriately in their own work.

JEL Classification: C18, C8, C810, J00, J01, I2

Keywords: current population survey, education, monthly outgoing rotation group, National Bureau of Economic Research data, missing data, coding errors, NBER

Corresponding author:

Joni Hersch
Vanderbilt Law School
131 21st Avenue South
Nashville, TN 37203
USA

E-mail: joni.hersch@vanderbilt.edu

1. Introduction

In 2015 the Bureau of Labor Statistics (BLS) and the United States Census Bureau modified questions related to educational attainment in the Current Population Survey (CPS). In particular, a question regarding the length of a person's master's program was eliminated, as well as questions that asked people with bachelor's degrees if they had ever taken graduate or professional courses, and if they had completed six or more of those courses.

We have found that this change in the CPS has generated effects that should be addressed when using data obtained from the CPS Merged Outgoing Rotation Groups (MORG), compiled by the National Bureau of Economic Research (NBER).¹ As part of course activities in two classes in which we used the NBER MORG extract,² we discovered that the revision to the CPS affected coding used by the NBER to calculate their variable, "Imputed Highest Grade Completed" (*ihigrdc*). Imputed Highest Grade Completed is based on the credential oriented CPS variable *grade92* (*peeduca* in the original CPS data), that is coded from 31 to 46. Specifically, the CPS variable collapses into four categories those with less than 9th grade education, and codes degrees without corresponding years of schooling for those with more than 11th grade education. Thus, the imputed years of education provided in the NBER MORG extract provides a convenient measure of years of education. Because the NBER coding continues to use prior coding to calculate this variable, the omission of these questions from the CPS results in a substantial number of missing values for *ihigrdc*.

Upon discovering this problem, we notified the NBER on January 25, 2019 of the large number of missing values for *ihigrdc* and described what we believed to be the source of the

¹ "CPS Merged Outgoing Rotation Groups," *National Bureau of Economic Research*, last accessed July 12, 2019, <https://www.nber.org/data/morg.html> [<https://perma.cc/QX7E-ULSZ>].

² Vanderbilt Law School Ph.D. courses: Econometrics for Legal Research; Labor Markets and Human Resources II.

missing values. An exchange of emails on January 28, 2019 appeared to confirm the source of our concern. However, as of this writing (July 15, 2019), the 2015–2017 MORG extract using the unmodified coding remains on the NBER website. The 2018 extract, which was posted after January 28, 2019 and also uses unmodified coding, also remains on the NBER website.

The problem is readily apparent in Table 1. When observing the tabulated data for the *ihigrdc* variable between 2011 and 2018, 27.5% of the data has a missing value recorded between the years of 2015 and 2018. In sharp contrast, none of the data has a missing value recorded between 2011 and 2014. Upon further inspection, we find that the entirety of the 2015–2018 missing values corresponds to individual bachelor’s degrees (e.g., BA, AB, BS) and master’s degrees (e.g., MA, MS, Meng, Med, MSW, MBA). One immediate consequence of using these data is a substantial underestimation of years of education if based on *ihigrdc*. The error results in an underestimation of approximately 1.2 years of education when using MORG data after 2014.

2. CPS Questions

In January 2015, the BLS and the U.S. Census Bureau modified the CPS to add questions on certifications and licenses.³ As part of this process, they eliminated other questions to ensure the survey was not more cumbersome for respondents. Three educational attainment questions that had been added in 1996 were eliminated after the two bureaus determined that “[d]ata from these questions had never been published by BLS, and a literature review found very few publications using the questions.” The eliminated questions are the following:

1. Since completing your bachelor’s degree, have you ever taken any graduate or professional school courses for credit? [Asked of people whose

³ Mary Dorinda Allard, “Adding Questions on Certifications and Licenses to the Current Population Survey,” *Monthly Labor Review*, U.S. Bureau of Labor Statistics, last accessed July 1, 2018, <https://doi.org/10.21916/mlr.2016.52> [<https://perma.cc/TB87-HL8N>].

- highest educational attainment was a bachelor's degree] (*grprof* in the original CPS data, *pegrprof* in the MORG data, see Table 2)
2. Did you complete six or more graduate or professional school courses? [Asked of people who responded "yes" to the previous question] (*gr6cor* in the original CPS data, *pegr6cor* in the MORG data, see Table 3)
 3. Was your master's degree program a 1-year, 2-year, or 3-year program? [Asked of people whose highest educational attainment was a master's degree] (*pems123* in the original CPS data, *ms123* in the MORG data, see Table 4)

3. Imputing Highest Grade Completed

In the NBER's data dictionary, the NBER imputes the variable *ihigrdc*, following the method described in Jaeger (2002).⁴ The NBER has offered this variable since 1998 in order to have a comparable measure to the "highest grade completed" variable available before 1992. The NBER's current method, which uses the questions that were eliminated in 2015, was intended to add precision to the years of education measure for those who have bachelor's or master's degrees.

The *ihigrdc* measure takes on values within the 0–18 range, with professional and doctoral degrees receiving a value of 18. Jaeger (2002)'s method establishes that "[z]ero through six post-B.A. courses are assumed to constitute one year and are assigned a highest grade completed of 17." Individuals who took more than six post-bachelor's courses are assigned a value of 18. Furthermore, individuals with one-year master's degree programs are assigned 17, and those from 2- or 3-year programs are assigned a value of 18.

A review of the NBER code to generate *ihigrdc* reveals its reliance on the existence of the variables *gr6cor* and *ms123* in order to generate the values 17 and 18.⁵ Because *gr6cor* and

⁴ Daniel Feenberg and Jean Roth, "CPS Labor Extracts 1979–2006," *National Bureau of Economic Research*, last accessed July 1, 2018, <https://www.nber.org/morg/docs/cpsx.pdf> [<https://perma.cc/75X8-42ZX>].

⁵ NBER MORG documentation, last accessed July 15, 2018, <https://www.nber.org/morg/sources/annual2015.log> [<https://perma.cc/6KRT-P93P>], <https://www.nber.org/morg/sources/annual2016.log> [<https://perma.cc/7E3H-8B4W>],

ms123 have not been available since 2014, the code leads 27.5% of observations, all of which should have been coded as 17 or 18, to have missing values.

The next question is how to correct for this issue. The goal of Jaeger (2002) regarding these questions was to obtain a more accurate measure for educational attainment for those with post-graduate studies. For the March 1998 sample, he shows that the average highest grade completed is 0.11 grades higher when using his method, as opposed to using only the “highest degree received” question without taking into account any post-BA course work that does not result in an additional degree.

In order to ascertain the magnitude of the omission, we first replicate the method used by the NBER to calculate *ihigrdc* as a baseline. This replication can be found in Tables 5 through 8, under the columns headed as “MORG.” We have subsequently generated a modified version of the code, in which we have changed only the imputation for observations relying on the eliminated questions, which we have labeled as “Modified.”⁶

We first set out to see the variation caused by the change in method using 2011 to 2014 data in order to establish how much change is attributable to the method change. For this period, and considering all observations, Jaeger’s method allows us to find an average highest grade completed of 13.438 years, while our modified method gives us an average of 13.280 years. The difference ranges between 0.153 and 0.160 in these years, as can be seen in Table 5. This difference increases to 0.184 (ranging between 0.181 and 0.187), when restricting the data to

<https://www.nber.org/morg/sources/annual2017.log> [<https://perma.cc/657Y-VXES>].
<https://www.nber.org/morg/sources/annual2018.do>.

⁶ Considering that Jaeger (2002) makes use of other educational attainment questions that continue to be available after the 2015 changes, our modified calculation remains identical to the MORG data for all questions that have remained unchanged. This leads to no changes in the imputed highest grade completed for observations with *ihigrdc* values ranging between 0–15. Therefore, all changes can be found amongst those within the 16–18 range. Specifically, bachelor’s degree is assigned 16 years; master’s degree is assigned 17 years; professional school and doctorate degrees are assigned 18 years. The complete coding we used to create our “Modified” variable is provided in Figure 1.

individuals between the ages of 25 and 64 (Table 6). While this number is greater than the 0.11 that Jaeger (2002) finds when implementing his method, our modified method still partially follows his method, when possible, particularly with respect to the attainment of high school diplomas through traditional methods or by way of a GED. Considering that this adjustment for GED-obtained degrees will have a negative effect on the imputed highest completed degree variable, our calculations appear reasonable.

Focusing on the period of interest, 2015 to 2018, we find that the differences are understandably greater. As shown in Table 7, the omission of 27.5% of the observations that were at the higher end of the spectrum translates to an underestimation of highest grade completed by 1.091 grades on average—ranging between 1.065 and 1.115 on a per-year basis. Restricting for ages 25 to 64 once again leads to a larger difference, with an average of 1.208 grades for the 2015 through 2018 period (Table 8). Considering that our modified method results in a highest completed degree variable with a lower average than the one generated by the NBER method (as shown in Tables 5 and 6), the average underestimation would likely be between 1.37 and 1.39 years of education, or approximately 10% of the actual value.

4. Conclusion

Accounting for the 2015 changes to the educational attainment questions in the CPS—particularly the questions related to post-graduate studies for individuals with bachelor’s degrees—is critical. Because these changes remain unaccounted for by the NBER in their widely-used CPS MORG data, 27.5% of their observations contain missing values for their Imputed Highest Grade Completed variable, *ihigrdc*. These missing values, in turn, lead to an average Imputed Highest Grade Completed of about 1.2 fewer years than would result from

using the whole data set. Ultimately, our intent is to make users of these data aware of this crucial issue, so that they may address the issue appropriately in their own work.

Table 1: Number of Observations Corresponding to Each Imputed Highest Grade of School Completed (*ihigrdc*) 2011–2018

<i>ihigrdc</i>	2011	2012	2013	2014	2015	2016	2017	2018
0	1,083	1,102	968	971	1,009	1,033	930	953
2.5	2,107	2,093	1,954	2,043	2,075	1,974	1,807	1,724
5.5	4,024	3,894	3,729	3,715	3,706	3,625	3,384	3,237
7.5	6,986	6,456	6,224	6,177	5,920	5,741	5,426	4,956
9	8,129	7,665	7,493	7,384	7,394	7,108	6,709	6,368
10	13,087	12,577	12,046	11,810	11,734	11,526	11,026	10,239
11	15,725	15,258	14,584	14,551	14,360	13,639	13,345	12,707
12	103,348	100,996	100,120	100,751	98,937	97,934	95,287	92,690
13	22,284	22,425	22,037	22,059	21,315	21,359	20,412	19,454
14	39,074	38,808	39,480	39,311	39,826	40,470	40,608	39,860
15	9,840	10,025	9,857	9,846	9,798	9,800	9,501	9,171
16	45,508	46,507	47,940	49,063	5,789	5,838	5,722	5,553
17	6,483	6,541	6,568	6,523	0	0	0	0
18	40,656	42,061	42,717	42,852	9,196	9,304	9,272	9,358
(missing)	0	0	0	0	83,633	85,522	86,560	86,062
Total	318,334	316,408	315,717	317,056	314,692	314,873	309,989	302,332

Source: NBER MORG Data 2011–2018.

Note: The columns under each year heading report the number of observations in each cell.

Table 2: Completed any Graduate or Professional School Courses 2011–2018

grprof	2011	2012	2013	2014	2015	2016	2017	2018
1 (yes)	16,501	17,098	16,804	16,164	0	0	0	0
2 (no)	40,154	40,984	42,073	43,032	0	0	0	0
(missing)	261,679	258,326	256,840	257,860	314,692	314,873	309,989	302,332
Total	318,334	316,408	315,717	317,056	314,692	314,873	309,989	302,332

Source: NBER MORG Data 2011–2018.

Note: The columns under each year heading report the number of observations in each cell.

Table 3: Completed 6 or more Graduate or Professional School Courses 2011–2018

gr6cor	2011	2012	2013	2014	2015	2016	2017	2018
1 (yes)	11,833	12,441	12,131	11,569	0	0	0	0
2 (no)	4,668	4,657	4,673	4,595	0	0	0	0
(missing)	301,833	299,310	298,913	300,892	314,692	314,873	309,989	302,332
Total	318,334	316,408	315,717	317,056	314,692	314,873	309,989	302,332

Source: NBER MORG Data 2011–2018.

Note: The columns under each year heading report the number of observations in each cell.

Table 4: Duration of Master's Degree Programs 2011–2018

ms123	2011	2012	2013	2014	2015	2016	2017	2018
1	1,815	1,884	1,895	1,928	0	0	0	0
2	14,207	14,662	15,113	15,431	0	0	0	0
3	6,189	6,234	6,611	6,766	0	0	0	0
(missing)	296,123	293,628	292,098	292,931	314,692	314,873	309,989	302,332
Total	318,334	316,408	315,717	317,056	314,692	314,873	309,989	302,332

Source: NBER MORG Data 2011–2018.

Note: The columns under each year heading report the number of observations in each cell. Values of *ms123* in years.

Table 5: Distribution of Imputed Highest Grade Completed 2011–2014

Imputed Highest Grade Completed	2011		2012		2013		2014		All	
	MORG	Modified								
0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.5	0.007	0.007	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.006
5.5	0.013	0.013	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
7.5	0.022	0.022	0.020	0.020	0.020	0.020	0.019	0.019	0.020	0.020
9	0.026	0.026	0.024	0.024	0.024	0.024	0.023	0.023	0.024	0.024
10	0.041	0.041	0.040	0.040	0.038	0.038	0.037	0.037	0.039	0.039
11	0.049	0.049	0.048	0.048	0.046	0.046	0.046	0.046	0.047	0.047
12	0.325	0.325	0.319	0.319	0.317	0.317	0.318	0.318	0.320	0.320
13	0.070	0.070	0.071	0.071	0.070	0.070	0.070	0.070	0.070	0.070
14	0.123	0.123	0.123	0.123	0.125	0.125	0.124	0.124	0.124	0.124
15	0.031	0.031	0.032	0.032	0.031	0.031	0.031	0.031	0.031	0.031
16	<i>0.143</i>	<i>0.195</i>	<i>0.147</i>	<i>0.201</i>	<i>0.152</i>	<i>0.205</i>	<i>0.155</i>	<i>0.206</i>	<i>0.149</i>	<i>0.202</i>
17	<i>0.020</i>	<i>0.070</i>	<i>0.021</i>	<i>0.072</i>	<i>0.021</i>	<i>0.075</i>	<i>0.021</i>	<i>0.076</i>	<i>0.021</i>	<i>0.073</i>
18	<i>0.128</i>	<i>0.026</i>	<i>0.133</i>	<i>0.028</i>	<i>0.135</i>	<i>0.028</i>	<i>0.135</i>	<i>0.029</i>	<i>0.133</i>	<i>0.028</i>
Mean	13.355	13.202	13.423	13.264	13.482	13.322	13.491	13.334	13.438	13.280
Mean Difference	0.153		0.159		0.160		0.157		0.158	
N	318,334		316,408		315,717		317,056		1,267,515	

Source: Calculations from weighted 2011–2014 NBER MORG Data.

Note: “MORG” refers to the variable *ihigrdc* from the original data set. “Modified” refers to the modified imputed value that does not use the educational attainment questions excluded since 2015.

Table 6: Distribution of Imputed Highest Grade Completed 2011–2014, Ages 25–64

Imputed Highest Grade Completed	2011		2012		2013		2014		All	
	MORG	Modified								
0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.5	0.006	0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006	0.006
5.5	0.013	0.013	0.013	0.013	0.012	0.012	0.012	0.012	0.012	0.012
7.5	0.015	0.015	0.015	0.015	0.014	0.014	0.014	0.014	0.014	0.014
9	0.018	0.018	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
10	0.024	0.024	0.024	0.024	0.023	0.023	0.022	0.022	0.023	0.023
11	0.032	0.032	0.031	0.031	0.030	0.030	0.030	0.030	0.031	0.031
12	0.311	0.311	0.305	0.305	0.303	0.303	0.303	0.303	0.306	0.306
13	0.068	0.068	0.069	0.069	0.067	0.067	0.067	0.067	0.068	0.068
14	0.133	0.133	0.133	0.133	0.135	0.135	0.133	0.133	0.133	0.133
15	0.030	0.030	0.030	0.030	0.029	0.029	0.029	0.029	0.030	0.030
16	<i>0.172</i>	<i>0.232</i>	<i>0.176</i>	<i>0.238</i>	<i>0.181</i>	<i>0.242</i>	<i>0.185</i>	<i>0.242</i>	<i>0.178</i>	<i>0.238</i>
17	<i>0.023</i>	<i>0.084</i>	<i>0.023</i>	<i>0.086</i>	<i>0.023</i>	<i>0.088</i>	<i>0.023</i>	<i>0.090</i>	<i>0.023</i>	<i>0.087</i>
18	<i>0.152</i>	<i>0.031</i>	<i>0.156</i>	<i>0.031</i>	<i>0.157</i>	<i>0.032</i>	<i>0.156</i>	<i>0.032</i>	<i>0.155</i>	<i>0.031</i>
Mean	13.759	13.577	13.807	13.620	13.855	13.669	13.858	13.677	13.819	13.636
Mean Difference	0.182		0.187		0.186		0.181		0.184	
N	215,361		212,848		211,191		210,574		849,974	

Source: Calculations from weighted 2011–2014 NBER MORG Data.

Note: Individuals 25–64 years old. “MORG” refers to the variable *ihigrdc* from the original data set. “Modified” refers to the modified imputed value that does not use the educational attainment questions excluded since 2015.

Table 7: Distribution of Imputed Highest Grade Completed 2015–2018

Imputed Highest Grade Completed	2015		2016		2017		2018		All	
	MORG	Modified								
0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.5	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
5.5	0.012	0.012	0.012	0.012	0.011	0.011	0.011	0.011	0.011	0.011
7.5	0.019	0.019	0.018	0.018	0.018	0.018	0.016	0.016	0.018	0.018
9	0.023	0.023	0.023	0.023	0.022	0.022	0.021	0.021	0.022	0.022
10	0.037	0.037	0.037	0.037	0.036	0.036	0.034	0.034	0.036	0.036
11	0.046	0.046	0.043	0.043	0.043	0.043	0.042	0.042	0.044	0.044
12	0.314	0.314	0.311	0.311	0.307	0.307	0.307	0.307	0.310	0.310
13	0.068	0.068	0.068	0.068	0.066	0.066	0.064	0.064	0.066	0.066
14	0.127	0.127	0.129	0.129	0.131	0.131	0.132	0.132	0.129	0.129
15	0.031	0.031	0.031	0.031	0.031	0.031	0.030	0.030	0.031	0.031
16	<i>0.018</i>	<i>0.207</i>	<i>0.019</i>	<i>0.209</i>	<i>0.018</i>	<i>0.216</i>	<i>0.018</i>	<i>0.219</i>	<i>0.018</i>	<i>0.213</i>
17	<i>0.000</i>	<i>0.078</i>	<i>0.000</i>	<i>0.081</i>	<i>0.000</i>	<i>0.082</i>	<i>0.000</i>	<i>0.084</i>	<i>0.000</i>	<i>0.081</i>
18	<i>0.029</i>	<i>0.029</i>	<i>0.030</i>	<i>0.030</i>	<i>0.030</i>	<i>0.030</i>	<i>0.031</i>	<i>0.031</i>	<i>0.030</i>	<i>0.030</i>
(missing)	0.266	0.000	0.272	0.000	0.279	0.000	0.285	0.000	0.275	0.000
Mean	12.286	13.351	12.315	13.397	12.351	13.452	12.377	13.492	12.332	13.422
Mean Difference		-1.065		-1.082		-1.101		-1.115		-1.091
N	314,692		314,873		309,989		302,332		1,241,886	

Source: Calculations from weighted 2015–2018 NBER MORG Data.

Note: “MORG” refers to the variable *ihigrdc* from the original data set. “Modified” refers to the modified imputed value that does not use the educational attainment questions excluded since 2015.

Table 8: Distribution of Imputed Highest Grade Completed 2015–2018, Ages 25–64

Imputed Highest Grade Completed	2015		2016		2017		2018		All	
	MORG	Modified								
0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.5	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005
5.5	0.012	0.012	0.012	0.012	0.011	0.011	0.011	0.011	0.012	0.012
7.5	0.015	0.015	0.014	0.014	0.013	0.013	0.013	0.013	0.014	0.014
9	0.017	0.017	0.017	0.017	0.016	0.016	0.016	0.016	0.016	0.016
10	0.022	0.022	0.023	0.023	0.022	0.022	0.020	0.020	0.022	0.022
11	0.031	0.031	0.028	0.028	0.029	0.029	0.028	0.028	0.029	0.029
12	0.298	0.298	0.296	0.296	0.293	0.293	0.290	0.290	0.294	0.294
13	0.065	0.065	0.065	0.065	0.062	0.062	0.061	0.061	0.063	0.063
14	0.135	0.135	0.137	0.137	0.139	0.139	0.139	0.139	0.138	0.138
15	0.030	0.030	0.030	0.030	0.029	0.029	0.028	0.028	0.029	0.029
16	0.020	0.242	0.021	0.244	0.020	0.250	0.020	0.255	0.020	0.247
17	0.000	0.092	0.000	0.095	0.000	0.095	0.000	0.097	0.000	0.095
18	0.033	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.033	0.033
(missing)	0.314	0.000	0.318	0.000	0.324	0.000	0.332	0.000	0.322	0.000
Mean	12.496	13.687	12.526	13.725	12.560	13.771	12.582	13.814	12.540	13.748
Mean Difference		-1.191		-1.199		-1.211		-1.232		-1.208
N	207,681		206,883		201,995		195,514		812,073	

Source: Calculations from weighted 2015–2018 NBER MORG Data.

Note: Individuals 25–64 years old. “MORG” refers to the variable *ihigrdc* from the original data set. “Modified” refers to the modified imputed value that does not use the educational attainment questions excluded since 2015.

Figure 1: Comparison between NBER and Modified Stata Code

NBER MORG code	Modified code
<pre> generate double ihigrdc = . ** Jaeger label var ihigrdc "Imputed highest grade completed" replace ihigrdc = 0 if grade92==31 replace ihigrdc = 2.5 if grade92==32 replace ihigrdc = 5.5 if grade92==33 replace ihigrdc = 7.5 if grade92==34 replace ihigrdc = 9 if grade92==35 replace ihigrdc =10 if grade92==36 replace ihigrdc =11 if grade92==37 replace ihigrdc =12 if grade92==38 replace ihigrdc = 0 if grade92==39 & ged==2 & gedhigr==1 replace ihigrdc = 2.5 if grade92==39 & ged==2 & gedhigr==2 replace ihigrdc = 5.5 if grade92==39 & ged==2 & gedhigr==3 replace ihigrdc = 7.5 if grade92==39 & ged==2 & gedhigr==4 replace ihigrdc = 9 if grade92==39 & ged==2 & gedhigr==5 replace ihigrdc = 10 if grade92==39 & ged==2 & gedhigr==6 replace ihigrdc = 11 if grade92==39 & ged==2 & gedhigr==7 replace ihigrdc = 12 if grade92==39 & ged==2 & gedhigr==8 replace ihigrdc = 12 if ged==1 replace ihigrdc = 12 if grade92>=40 & grade92<=42 & yrcoll==1 replace ihigrdc = 13 if grade92>=40 & grade92<=42 & yrcoll==2 replace ihigrdc = 14 if grade92>=40 & grade92<=42 & yrcoll==3 replace ihigrdc = 15 if grade92>=40 & grade92<=42 & yrcoll==4 replace ihigrdc = 16 if grade92>=40 & grade92<=42 & yrcoll==5 replace ihigrdc = 16 if grade92==43 & grprof==2 replace ihigrdc = 17 if grade92==43 & gr6cor==2 replace ihigrdc = 18 if grade92==43 & gr6cor==1 replace ihigrdc = 17 if grade92==44 & ms123==1 replace ihigrdc = 18 if grade92==44 & ms123>=2 & ms123<. replace ihigrdc = 18 if grade92==45 grade92==46 </pre>	<pre> generate double Modified = . replace Modified = ihigrdc replace Modified = 16 if grade92==43 replace Modified = 17 if grade92==44 </pre>

Source: NBER Code is obtained from the NBER MORG documentation, last accessed July 15, 2018, <https://www.nber.org/morg/sources/annual2015.log> [<https://perma.cc/6KRT-P93P>], <https://www.nber.org/morg/sources/annual2016.log> [<https://perma.cc/7E3H-8B4W>], <https://www.nber.org/morg/sources/annual2017.log> [<https://perma.cc/657Y-VXES>], <https://www.nber.org/morg/sources/annual2018.do>.