A Data Appendix

This appendix provides detailed information on the data used in the analysis. In particular, sections A.1 to A.4 clarify how variables used in the analysis were constructed, and highlight any differences across the data sets.

A.1 Earnings

The NHIS provides information on total earnings in the last year divided into 11 categories.\(^1\) Taking the midpoint of the range to get a continuous measure of earnings, I then convert the continuous measure into hourly earnings using the reported number of hours that the individual worked in the past week.\(^2\) Because the NHIS and HRS data span several years, I use a BLS consumer price index to convert these earnings data into real 2004 dollars.

The measure of earnings constructed from the NIS is the individual’s self-reported current salary or wage. Individuals can choose to report their salaries or wages in various units including hourly, monthly and yearly. I convert the salaries into hourly salaries using information on hours usually worked per week. For the NIS results in Table 10 that add pre-immigration wage information of the last job abroad, the pre-immigration data are converted into real 2004 local currency using the Penn World Tables, and then converted in the 2004 U.S. dollars using OANDA exchange rate data. One extreme outlier of annual earnings over $1 million is dropped.

Respondents in the HRS can report their current salary in various units including hourly, biweekly, monthly and yearly. Given that many of the respondents are retired, I construct a pseudo-panel in this data using salary information in the most recent job if the individual is retired. The pseudo panel also includes information on the starting and ending salary in one additional long-term job prior to the current or most recent job for all respondents. I convert information on salaries reported per two weeks, per month and per year into hourly earnings using the corresponding hours per week worked in each of the jobs. Two extreme outliers of hourly earnings of over $3000 are dropped. Information about the industries and occupations associated with these previous jobs is asked, and the respondent’s age is adjusted.

\(^1\) The first range is 0 to $4999, and the second to last range is $65,000-$74,999.
\(^2\) This assumes that the number of hours worked in the past week is a good approximation of the number of hours the individual worked in the past year. However, the assumption is supported by the robustness of the results to the use of annual earnings rather than hourly earnings. These results are available on request.
appropriately for past labor market experiences. In the HRS pseudo-panel, the median year of employment data is 1986 and the earliest year of data is 1938.\(^3\)

In contrast to the other data sets, the key disadvantage of the HSE data is that income is not reported at the individual level. For the HSE data, I construct an individual level measure using joint annual income reported at the couple level. First, I convert 31 categories of joint annual income into a continuous measure that is based on the midpoint of the range of the category.\(^4\) After that, I transform the continuous measure of the couple’s income into an individual measure. In the majority of cases, the assignment is simple for the households where an individual is not married or is the only person in the household working. In other cases, the individuals’ share of joint income is weighted by whether they work full-time or part-time. For example, if both members are working full-time, the individual measure of income evenly divides their joint income. If one member works full-time and the other part-time, the member who works full-time is assigned three-quarters of the joint income and the remaining one-quarter is assigned to the part-time worker.\(^5\) The measure of income in the HSE is converted into 2004 pounds using a GDP deflator from the U.K. Office of National Statistics. No information about hours worked at the individual level is provided so the measure of real earnings used in the analysis for the HSE sample is the real annual earnings.

A.2 Height

Height is measured in centimeters by trained enumerators in the HSE. Height is self-reported in the NHIS, the NIS, and the 1992 wave of the HRS. The unit of measurement for height is inches in the NHIS and feet and inches in the HRS. Respondents of the NIS can choose whether to report their height in a combination of feet, inches, meters and centimeters. Without loss of information, I convert height in the HSE and NIS to inches for comparability across all data sets. I drop extreme outliers of height below 110 centimeters. This corresponds to a loss of three observations centimeters in the HSE and 24 observations in the NIS.\(^6\) This has no effect on the samples in the HRS and the NHIS.

---

\(^3\) To address concern regarding recall bias in past wages, I examined all of the results with only recent information on current job and the most recent job for retirees. The results are robust to this truncation and available upon request.

\(^4\) These category divisions are quite fine. The first category is under £520, the second category is from £520 to under £2600 pounds, and the second to last category is £140,000 to under £150,000.

\(^5\) The results obtained from this method are generally similar to the estimates over a sample of individuals who are the sole earners in their marriage.

\(^6\) The fact that there are more outliers in the NIS is not surprising given that respondents provide their height measurements and select the appropriate unit, and may have chosen the wrong unit to accompany the quantity.
A.3 Industry and Occupation

Many of the regression results presented in the paper include controls for industry and occupation. The specificity of the controls varies across data sets and depends on the level of detail available in each data set. The NHIS offers two-digit industry and occupation information. The method of classification of industries and occupations changed between 2004 and 2005 such that the waves 2000 to 2004 use one set of codes and 2005 to 2007 use a different set of codes. No bridge is offered between the systems of classification, so I allow the estimation to provide separate coefficients on each industry and occupation code in the two systems. The HSE records industry at the two-digit level and occupation at the three-digit level. In the HRS, only one-digit industry and occupation codes are available in the data. Finally, the NIS contains four-digit industry and occupation information but there are not enough observations to estimate this many fixed effects so I aggregate the information to the next level.

A.4 Region

The regression results presented in the paper include controls for region of residence in the U.S. or the U.K. In the NHIS, region of residence is comprised of four categories: northeast, midwest, south and west. The HSE reports the government office region in which the respondent resides: North East, North West and Merseyside, Yorkshire and Humberside, West Midlands, East Midlands, Eastern, London, South East and South West. The observations in the HRS are divided into nine regions that are recoded from the state of residence, and information regarding the specific states included in each of the nine codes is not provided. The NIS provides information on the state of residence, aggregated into fifteen categories.

The NIS and HSE include information on country and region of birth of immigrants. The categories in the NIS are Canada, China, Colombia, Cuba, Dominican Republic, El Salvador, Ethiopia, Guatemala, Haiti, India, Jamaica, Korea, Mexico, Nigeria, Peru, Philippines, Poland, Russia, Ukraine, U.K., Vietnam, Europe and Central Asia, South Asia, Other North America, Latin America and the Caribbean, Sub-Saharan Africa, Middle East and North Africa, Oceania and the Artic Region. The categories of the HSE are Ireland, West Indies, India, Pakistan, Kenya, Uganda, Tanzania, Hong Kong, China, Malaysia, Vietnam, Taiwan, Singapore, Other Africa, and other. The NHIS only has region of birth of immigrants; there are ten categories: Central America and the Caribbean, South America, Europe, Russia and former USSR areas, Africa, Middle East, Indian subcontinent, Asia, Southeast Asia, and elsewhere. The HRS does not provide any information about place of origin of immigrants.
A.5  Job Characteristics

Measures of job characteristics were created using the revised fourth edition (1991) of the Dictionary of Occupational Titles (DOT). In this paper, I consider the measures of physical strength and cognitive reasoning associated with each occupation. This strength measure takes on five values: sedentary (1), light (2), medium (3), heavy (4) and very heavy (5). The reasoning measure ranges from 1 to 6 where 1 involves one or two step instructions to carry out and 6 involves the complex analysis. I use crosswalks from the National Crosswalk Service Center to link the DOT data with occupation codes used household surveys. The HRS uses 1980 Census codes, and the NIS 2003 Census codes. Prior to 2004, the NHIS uses the 1980 Simple Occupational Classification (SOC) codes. The NHIS occupational coding system switches in 2005 and cannot be mapped into DOT. The HSE uses 2000 British SOC codes, and to my knowledge, there is no crosswalk between this system and the DOT codes.

Given the mapping across occupation codes, it was necessary to aggregate several jobs classified in the DOT to create measures of physical demands and reasoning corresponding to the data in the NHIS, HRS and NIS. This aggregation is a simple average of the DOT jobs associated with each four digit code to corresponding code in the three data sets. In the NHIS, construction laborers have one of the the highest level of physical demands at 3.50 while the lowest are associated with secretaries and workers in financial records processing with 1.11 and 1.09, respectively. Engineering requires a high level of reasoning at 5.24 with freight handlers on the other end at 1.66.

A.6  Sample Weights

The NIS, NHIS and HRS all include sample weights. My analysis with the HSE sample pools the 1999 and 2004 waves but the HSE did not calculate sample weights for the 1999 wave. The results reported in the paper do not use sample weights, but the regression results for the NIS, NHIS and HRS are all robust to the use of sample weights.

B  Controlling for Health and Cognition

Appendix Table 1 examines whether the inclusion of measures of health and cognitive ability reduces the wage returns to height for immigrants and natives. The regressions include health status in the NHIS and HSE samples and both health status and cognitive ability in the HRS sample. As expected given the positive correlation between height and health and height and cognition shown in Table 7, the returns to height decline with the inclusion of measures of health and cognition. However, the declines in the returns

7 The impact on the returns to height for natives and for immigrants is quite similar if indicators for health status and cognitive ability are included instead of as interval variables.
to height are quite small and the height premium in earnings remains substantially higher for male immigrants than for male natives. It is important to note that the height variable contains much more variation than self-reported health status which is divided coarsely into five categories. Similarly, the WAIS score available in the HRS is also fairly coarse with only 15 values. While the relationship between height and the measures of health and cognitive ability is stronger for immigrants than for natives, it is likely that there is residual variation in height that reflects differences in non-cognitive skills, cognitive ability or health beyond the coarse measures available in the data. Thus, the results of Appendix Table 1 can be consistent with the conclusion that the gap in the wage returns to height for male immigrants and natives is explained by differences in the mapping between height and health or height and cognitive and non-cognitive ability.

Appendix Table 1: Returns to Height Controlling for Health and Cognition

<table>
<thead>
<tr>
<th></th>
<th>NHIS Native (1)</th>
<th>NHIS Immig (2)</th>
<th>HSE Native (3)</th>
<th>HSE Immig (4)</th>
<th>HRS Native (5)</th>
<th>HRS Immig (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>0.005 [0.001]**</td>
<td>0.012 [0.002]**</td>
<td>0.009 [0.004]*</td>
<td>0.022 [0.008]**</td>
<td>0.008 [0.003]**</td>
<td>0.014 [0.008] +</td>
</tr>
<tr>
<td>Health Status</td>
<td>-0.062 [0.004]**</td>
<td>-0.048 [0.007]**</td>
<td>-0.131 [0.016]**</td>
<td>-0.094 [0.024]**</td>
<td>-0.021 [0.008]**</td>
<td>-0.109 [0.034]**</td>
</tr>
<tr>
<td>Cognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.007 [0.003]*</td>
<td>0.011 [0.012]</td>
</tr>
<tr>
<td>Observations</td>
<td>41537</td>
<td>9652</td>
<td>3519</td>
<td>1643</td>
<td>8098</td>
<td>722</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.32</td>
<td>0.40</td>
<td>0.33</td>
<td>0.34</td>
<td>0.16</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered by household in brackets. **, *, + denotes significance at the 1%, 5% and 10% level, respectively. The dependent variable is the logarithm of real wages. Height is measured in inches. All regressions include a quadratic in age, years of education, indicators for year, region, industry and occupation indicators, and a constant term.