Table 7: Intergenerational Earnings Elasticity Estimates Corrected for Measurement Error (With Percent Downward Bias Italicized)

	Parents	Children		
	(1)	**(2)**	**(3)**	**(4)**
Parent		0.372	0.326	
		(0.038)	(0.039)	
		52	*51*	
Grandparent	0.318		0.127	0.231
	(0.038)		(0.016)	(0.023)
	51		*27*	*50*

This table shows the estimated impact of measurement error on the multigenerational earnings elasticity coefficients. We numerically solve for the true intergenerational income elasticity given the observed biased coefficients and the parameters shown in Table 6. Standard errors were calculated as follows. First, the variance-covariance matrix of β_O was calculated by bootstrap with 1,000 replications. Next, the variance-covariance matrix for the numerical solution parameters were calculated by bootstrap with 100 replications from the CPS-DER. Finally, a bootstrap of the numerical solution was conducted with random draws from the β_O and input parameter distributions using the estimates as the means and variance-covariance matrices as the multivariate normal error distributions. The downward bias (in percent) is the percent difference between the error-corrected estimates and the coefficients in Table 5 Panel C. The estimated mobility difference between the two sets of generations (0.318 for grandparents to parents vs. 0.372 for parents to children) are entirely driven by changes in the variances as the estimated error-corrected correlation coefficients are 0.342 for both grandparents to parents (0.3415) and parents to children (0.3418).

Appendix Table 1: Intergenerational Mobility Regression Results for Male Lines
A. Years of Schooling

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.321	0.309			0.296
		(0.008)	(0.008)			(0.004)
Grandparent	0.346		0.026	0.132	0.351	
	(0.009)		(0.007)	(0.007)	(0.005)	
Observations	8,211				28,737	30,592

B. Income Elasticity (Earnings for Grandparents)

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.280	0.256			0.212
		(0.017)	(0.017)			(0.005)
Grandparent	0.177		0.072	0.117	0.167	
	(0.009)		(0.013)	(0.013)	(0.005)	
Observations	6,148				21,926	34,454

C. Earnings Elasticity

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.194	0.174			0.136
		(0.015)	(0.016)			(0.005)
Grandparent	0.175		0.076	0.116	0.162	
	(0.011)		(0.013)	(0.011)	(0.006)	
Observations	5,827				21,283	32,064

This table shows the results of four regressions for intergenerational mobility for several measures of status: A. years of schooling (for the most educated parent), B. log of family income (with earnings for grandparents as income is not available), and C. log of family earnings. In each case, the status measure is for families along the male lines in each generation. This includes the son's family, his father's family, and his paternal grandfather's family. The regression models include: 1) parents on grandparents, 2) children on parents, 3) children on parents and grandparents, and 4) children on grandparents on the sample where all three generations are observed and 5) parents on grandparents and 6) children on parents where the link to the third generation is not required. Each regression includes survey year fixed effects and controls for age and age-squared of the older partner in each generation.

Appendix Table 2: Intergenerational Mobility Regression Results for Female Lines

A. Years of Schooling

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.344	0.321			0.305
		(0.010)	(0.011)			(0.004)
Grandparent	0.280		0.035	0.125	0.274	
	(0.007)		(0.007)	(0.007)	(0.004)	
Observations	6,977				22,997	34,560

B. Income Elasticity (Earnings for Grandparents)

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.268	0.242			0.234
		(0.016)	(0.016)			(0.006)
Grandparent	0.180		0.099	0.42	0.168	
	(0.011)		(0.014)	(0.014)	(0.005)	
Observations	5,504				18,212	34,299

C. Earnings Elasticity

	Three Generation Sample				Two Generation Samples	
	Parents	Children			Parents	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Parent		0.167	0.150			0.149
		(0.015)	(0.015)			(0.005)
Grandparent	0.160		0.094	0.117	0.161	
	(0.013)		(0.014)	(0.014)	(0.006)	
Observations	5,094				17,565	31,713

This table shows the results of four regressions for intergenerational mobility for several measures of status: A. years of schooling (for the most educated parent), B. log of family income (with earnings for grandparents as income is not available), and C. log of family earnings. In each case, the status measure is for families along the female lines in each generation. This includes the daughter's family, her mother's family, and her maternal grandmother's family. The regression models include: 1) parents on grandparents, 2) children on parents, 3) children on parents and grandparents, and 4) children on grandparents on the sample where all three generations are observed and 5) parents on grandparents and 6) children on parents where the link to the third generation is not required. Each regression includes survey year fixed effects and controls for age and age-squared of the older partner in each generation.

Appendix A

Appendix A: Data Appendix

Person Identification Validation System

The Census Bureau uses the Person Identification Validation System (PVS) to assign unique PIKs to person records to facilitate unduplication and record linkage. The PVS uses a probabilistic matching algorithm and processes data through modules, with each module blocking the data in different ways and comparing different fields to find matches (Felligi and Sunter, 1969). To assign the PIKs, PVS compare personally identifiable information of the Personal identifiers are compared to data in a reference file constructed from the Social Security Administration (SSA) Numident file and other federal-agency administrative data, using different combinations of Social Security Numbers (SSNs), full name, full date of birth, and address (Wagner and Layne, 2014). Records cascade through matching modules, and only those observations that did not receive a PIK pass from one module to the next. All reference file records are available for linkage in each module. Through this process, the PVS appends PIKs to census, survey, and administrative records data.

PVS follows the typical steps in record linkage: preprocessing, sorting into blocks, identifying potential matches, and resolving best matches. Person records are preprocessed to standardize the blocking and matching fields between the census file and the reference file, ensuring that similar variables align. Next, PVS sorts the input and reference files into "blocks" or "cuts" for comparison. Blocking creates reasonably sized search spaces to find candidate matches, which is important with large files (Michelson and Knoblock, 2006). The SSA Numident contains nearly 500 million SSNs and forms the bulk of the reference file used in the matching process. The reference file is enhanced using administrative records to obtain additional variables not in the Numident, such as place of residence. The reference file includes all transactions associated with each SSN.[14] Consequently, the reference file is large and it is technically infeasible to compare every Numident record to every census record simultaneously. PVS compares census and reference file records within several passes in each module. Each pass alters the comparison of characteristics specified by the data processing staff, using slight variations of the match fields, with subsequent passes permitting more fuzziness in the match. For

[14] The SSA records transactions as corrections or name changes made on a particular SSN's record. The average number of transactions per SSN is 2.1 (Harris, 2014).

instance, the first pass may require that first name, last name, and year of birth match exactly, but the next pass may permit more tolerance in year of birth.

Within each pass of a module, potential matches are assigned a total score depending on the similarity of the characteristics of the input records and reference file records. PVS employs a string comparator program to measure Jaro-Winkler distances between first and last names in the input and reference files (Winkler, 1995). The PVS string comparator was developed by Winkler (1995) and measures the distance between two strings on a scale from 0 to 900, where a distance score of 0 is given if there is no similarity between two text strings and a score of 900 is given for an exact match. The cutoff value for the string distance is set to 750. These distances serve as a metric of how closely two names match, while allowing for some degree of misspelling. For numeric variables, such as year of birth, a maximum acceptable difference between the variable value in the input and reference record is programmable. This also allows for creation of an interval, or band, around year of birth to permit inexact matches. The total score is calculated as the sum of the agreement and disagreement weights attributed to each matching variable (Felligi and Sunter, 1969).

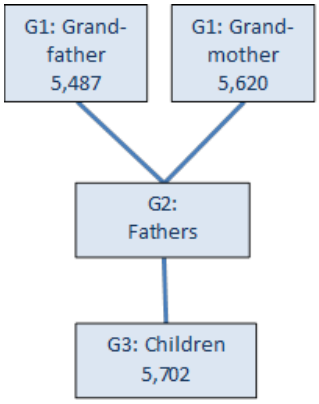
Potential matches are identified within each pass of a module, and only those with an overall score greater than a user-specified cutoff score are retained as potential matches. Input records that do not receive a match in the first pass move to the next pass. Once the input data has been processed through all passes of a module, potential matches are grouped into one file and sorted by person and by score.

The final step of a module evaluates the potential matches. The matches with the highest scores are processed using a decision rule to determine if a PIK will be assigned. If one potential match has a higher score than all the other potential matches for a particular input record, then the PIK associated with that reference record is assigned to that input observation. If there are multiple potential matches for a particular input observation with the same high score, then no PIK is assigned in that module. Records that fail to find a match in a module are passed along to the next module.

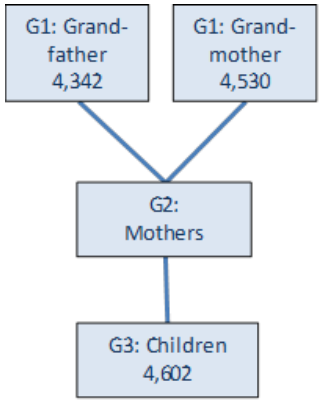
Building Multiple Generations

Appendix Figure 1: Three Generational Links

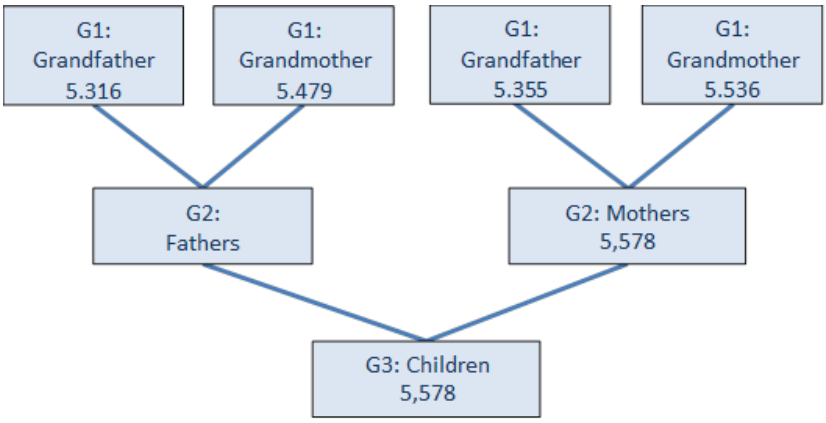
Panel A: Paternal Line Only



Panel B:Maternal Line Only



Panel C: Both Paternal Line and Maternal Line Present



Education Recode

1940 'Higrade' IPUMS code	1940 Code Label	Years of Schooling	Recode to match 2000	Consistent Years of Schooling Across Three Generations
0-4	less than 1st grade	0	1	0
4	1st grade	1	2	2
5	2nd grade	2	2	2
6	3rd grade	3	2	2
7	4th grade	4	2	2
8	5th grade	5	3	5.5
9	6th grade	6	3	5.5
10	7th grade	7	4	7.5
11	8th grade	8	4	7.5
12	9th grade	9	5	9
13	10th grade	10	6	10
14	11th grade	11	7	11
15	12th grade	12	9	12
16	1 year college	13	11	13
17	2 year college	14	12	14
18	3 year college	15	12	14
19	4 year college	16	13	16
20	5+ year college	17	14	17

Appendix B

Derivation of Biased Coefficients from Regression of Child Status on Parent and Grandparent Status

We can solve for the biased coefficients from the regression of child status on parent and grandparent status in a *relatively* simple form. For the parent coefficient ($\hat{\beta}_P$), it is then equal to:

$$\hat{\beta}_P = \frac{E(y_{t-2}^2)E(y_t y_{t-1}) - E(y_{t-1}y_{t-2})E(y_t y_{t-2})}{E(y_{t-1}^2)E(y_{t-2}^2) - E(y_{t-1}y_{t-2})^2} \tag{B.1}$$

Substituting (0.2) to replace each y_t with $(1+\delta)y_t^* + v_t$, we get:

$$\hat{\beta}_P = \frac{E[(1+\delta)^2 y_{t-2}^* + v_{t-2}]E((1+\delta)^2 y_t^* y_{t-1}^*) - E((1+\delta)^2 y_{t-1}^* y_{t-2}^*)E((1+\delta)^2 y_t^* y_{t-2}^*)}{E[(1+\delta)^2 y_{t-1}^* + v_{t-1}]E[(1+\delta)^2 y_{t-2}^* + v_{t-2}] - E((1+\delta)^2 y_{t-1}^* y_{t-2}^*)^2} \tag{B.2}$$

If we normalize lifetime income in each period to have mean zero, and combine terms, we get:

$$\hat{\beta}_P = \frac{(1+\delta)^4[\text{var}(y_{t-2}^*)\text{cov}(y_t^*, y_{t-1}^*) - \text{cov}(y_{t-1}^*, y_{t-2}^*)\text{cov}(y_t^*, y_{t-2}^*)] + (1+\delta)^2\text{var}(v_{t-2})\text{cov}(y_t^*, y_{t-1}^*)}{[(1+\delta)^2\text{var}(y_{t-1}^*) + \text{var}(v_{t-1})][(1+\delta)^2\text{var}(y_{t-2}^*) + \text{var}(v_{t-2})] - (1+\delta)^4\text{cov}(y_{t-1}^*, y_{t-2}^*)^2} \tag{B.3}$$

Without non-classical ($\delta = 0$) and classical ($\text{var}(v_t) = 0$ in each generation) measurement error, this reduces to:

$$\hat{\beta}_P = \frac{\text{var}(y_{t-2}^*)\text{cov}(y_t^*, y_{t-1}^*) - \text{cov}(y_{t-1}^*, y_{t-2}^*)\text{cov}(y_t^*, y_{t-2}^*)}{\text{var}(y_{t-1}^*)\text{var}(y_{t-2}^*) - \text{cov}(y_{t-1}^*, y_{t-2}^*)^2} \tag{B.4}$$

which is the unbiased parent coefficient. We could do the same for the grandparent coefficient ($\hat{\beta}_G$) to get:

$$\hat{\beta}_G = \frac{(1+\delta)^4[\text{var}(y_{t-1}^*)\text{cov}(y_t^*, y_{t-2}^*) - \text{cov}(y_{t-1}^*, y_{t-2}^*)\text{cov}(y_t^*, y_{t-1}^*)] + (1+\delta)^2\text{var}(v_{t-1})\text{cov}(y_t^*, y_{t-2}^*)}{[(1+\delta)^2\text{var}(y_{t-1}^*) + \text{var}(v_{t-1})][(1+\delta)^2\text{var}(y_{t-2}^*) + \text{var}(v_{t-2})] - (1+\delta)^4\text{cov}(y_{t-1}^*, y_{t-2}^*)^2} \tag{B.5}$$

which also reduces to the unbiased coefficient if $\delta = 0$ and $\text{var}(v_t) = 0$ in each generation.