A Mighty Toll:

Mine Accidents and the Long-Run Effect of Losing a Father Among Sons

Ezra G. Goldstein*

*A set of replication codes is available at https://doi.org/10.3886/E209584

Appendices

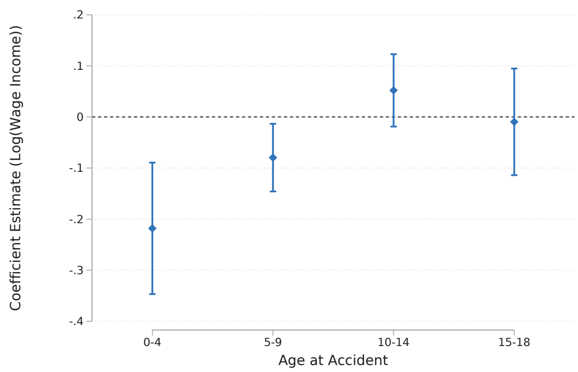
A Additional Figures and Tables

TABLE 4.—Fatal Accidents Inside and Outside of Mines

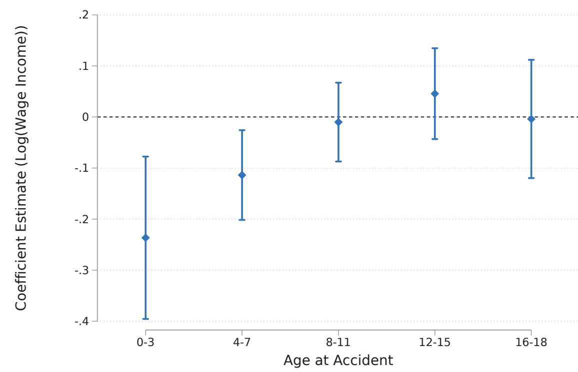
Date of accident	Name of Person	Nationality	Occupation	Age	Married or single	Number of widows	Number of orphans	Name of Mine	County	Nature and Cause of Accident in Brief
Jan. 20	Patrick McGuire, -----	Irish, -------	Driver, -------	24	M.	1	----	Colonial No. 4, ------		Fatally injured by being caught between mine cars on entry.
Feb. 6	Mike Verable, ---------	Slavonian,	Driver, -------	25	M.	1	----	Paul, ---------------		Instantly killed. Supposed to have fallen off car on entry and run over.
16	Lesto Kotona, -------	Magyar, ---	Miner, --------	34	M.	1	2	Adelaide, -----------		Killed by fall of roof while working on pillar stump.
17	John Malone, ---------	American, --	Elevator man,	24	M.	1	2	Colonial No. 4, ------		Fatally injured while trying to oil machinery in motion, contrary to orders. Outside.
March 3	George Shirrosky, ----	Slavonian,	Miner, -------	51	M.	*	5	Sumner No. 2, ------		Fatally injured by fall of coal at face of room while watching his fellow work-man pull the coal down.
4	Andrew Kolick, -------	Polish, ----	Miner, -------	22	S.	----	----	Trotter, -----------		Killed by fall of roof in pillar work while preparing to draw props. The gob slid knocking out posts.
June 17	Charles McGurke, ----	Irish, -------	Driver, -------	17	S.	----	----	Phillips, -----------	Fayette, ----	Fatally injured by being squeezed between car and rib on narrow side of entry.
July 26	Edward Hardin, -----	American, --	Driver, -------	30	S.	----	----	Lemont No. 1, ------		Instantly killed between car and rib on entry when car left the track.
Aug. 3	Mike Bikowski, -------	Polish, ----	Company-man	26	S.	----	----	Adelaide, -----------		Instantly killed by fall of slate on entry. He neglected to go to a place of safety.
18	John Meson, ----------	Slavonian,	Miner, -------	25	S.	----	----	Lincoln No. 1, ------		Instantly killed by fall of roof while drawing posts in pillar work.
31	John Franoz, --------	Polish, ----	Miner, -------	29	M.	1	1	Oliver No. 3, -------		Instantly killed by fall of slate when post was knocked out in pillar work.
Sept. 3	John Mitka, ----------	Polish, ----	Miner, -------	37	M.	1	6	Davidson, ----------		Fatally injured by fall of roof while drawing pillars.
28	Thomas Hutchenson, -	Scotch, -----	Timberman, -	37	M.	1	4	Colonial No. 1, -----		Instantly killed by fall of slate on entry while trying to secure it with cross-timbers.
21	Adolph Rottler, ------	German, ---	Miner, -------	51	M.	1	5	Davidson, ----------		Fatally injured by fall of slate on haul-age road.

Figure A.1: Mining Accident Record Example Page

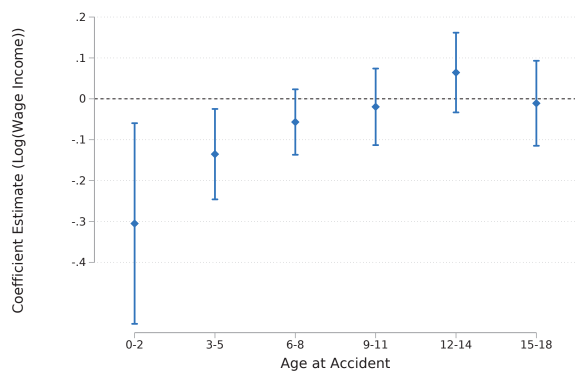
Notes: The figure presents a sample page of individual accident records published by state mining agencies. The example shows an excerpt of fatal accidents from Pennsylvania's Bituminous Coal Region during 1909 and highlights a typical record. Digitized accident records come from the Gerald E. Sherard Collection at the Colorado School of Mines and the Pennsylvania State Archives.



(a) 5-Year Groups



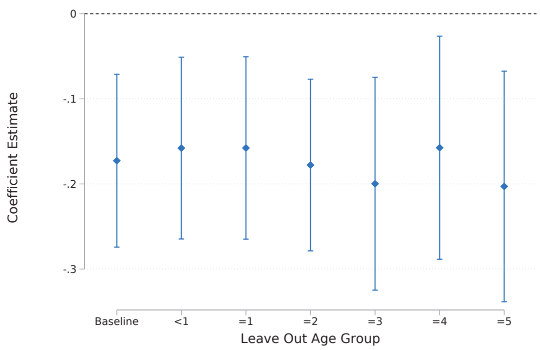
(b) 4-Year Groups



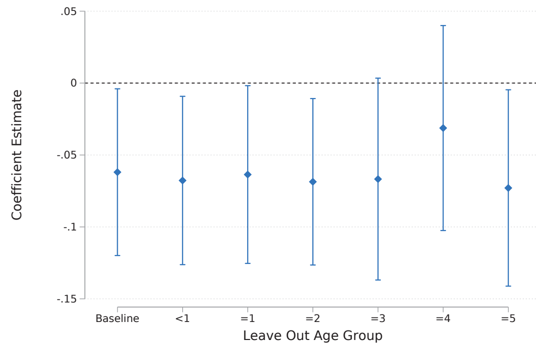
(c) 3-Year Groups

Figure A.2: Alternate Age Bin Sizes

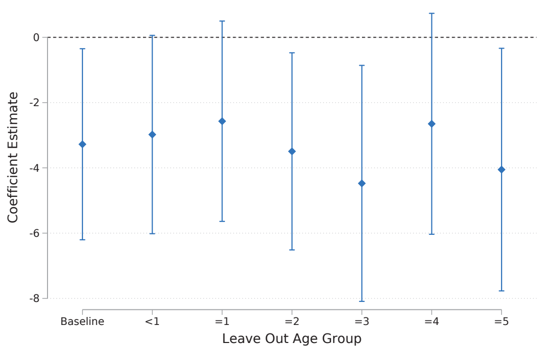
Notes: The figures plot the estimated effect of losing a father by age groups and across alternate group sizes. The figures plot the point estimates as solid points and 90 percent confidence intervals as lines. Figure A.2a presents a specification which places sons of accident victims in four groups by their age at the time of their father's accident. In Figures A.2b and A.2c, sons of accident victims are instead placed in five and six groups (each with progressively smaller bin sizes), respectively.



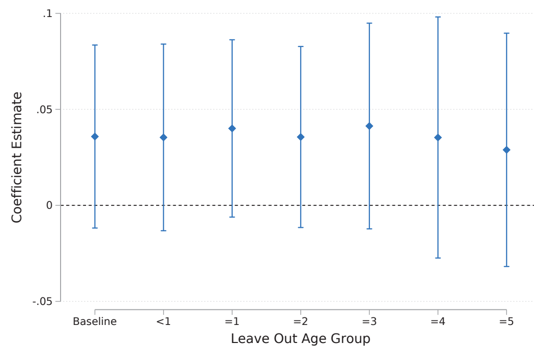
(a) Log(Wage Income)



(b) Above Median Income



(c) Income Percentile



(d) Non-Wage Income

Figure A.3: Robustness to Dropping Age Groups

Notes: This figure shows the robustness of the effect of parental loss at an early age on several measures of income. Panels A through D of the figure present the estimates (and 90 percent confidence intervals) for each measure, separately. In each panel, the leftmost estimate presents the baseline estimates reported in the main analysis. Each other estimate presents the results from omitting sons who were a specific age at the time of their father's accident.

Table A.1: Description of Accident Records

	Mean	Std. Dev.	Min	Max
Panel A: All Accident Records (N=183,536)				
Age	34.49	12.07	11	86
Birth Year	1880.76	13.09	1819	1916
Accident Year	1915.25	7.87	1900	1929
Fatal Accident	0.26	0.44	0	1
Pennsylvania	0.64	0.48	0	1
Illinois	0.23	0.42	0	1
West Virginia	0.11	0.31	0	1
Ohio	0.03	0.17	0	1
Panel B: Accident Causes (N=110,699)				
Fall of Rock	0.46	0.5	0	1
Cars and Machinery	0.35	0.48	0	1
Explosions and Gas	0.12	0.33	0	1
Other Causes	0.08	0.27	0	1

Notes: The table presents summary statistics for the state mining accident records. Panel A shows summary statistics for all records listed collected from the individual accident records of Pennsylvania, Illinois, West Virginia and Ohio. Panel B presents the three major causes of mining accidents listed in the Pennsylvania accident records, the only state for which causes were digitized. Fall of rock is any cause that mentions either fall of coal, roof, slate, or rock. Cars and Machinery contains any cause that faults runaway cars, crushed by cars, trapped by machinery, or faulty trucks or motors. Explosions and Gas contains any cause that lists dynamite explosions, faulty powder shots, delayed blasts, or dangerous gas. The remaining accidents are classified as Other and contain slips and falls (either by an individual or being hit by a tool), injuries from animals, struck by timber, suffocation due to poor ventilation, and electrocution. Digitized accident records come from the Gerald E. Sherard Collection at the Colorado School of Mines and the RG-45 Records of the Department of Mines and Mineral Industries from the Pennsylvania State Archives.

Table A.2: Summary Statistics of Fathers' Census Records

	(1) All Fathers	(2) Accident Fathers	(3) Fatal Fathers	(2)-(1) Differences	(3)-(1) Differences
Age	42.98	38.48	38.64	-4.320***	-4.121***
	[12.63]	[9.80]	[9.92]	(0.070)	(0.143)
Married	0.95	0.97	0.97	0.019***	0.012***
	[0.21]	[0.16]	[0.18]	(0.001)	(0.003)
Number of Children	2.79	3.18	3.22	0.378***	0.372***
	[1.80]	[1.96]	[1.95]	(0.014)	(0.028)
White	0.98	0.98	0.98	0.002**	-0.003
	[0.14]	[0.13]	[0.15]	(0.001)	(0.002)
Foreign Born	0.29	0.47	0.48	0.172***	0.194***
	[0.45]	[0.50]	[0.50]	(0.004)	(0.007)
Homeowner	0.45	0.36	0.34	-0.077***	-0.102***
	[0.50]	[0.48]	[0.47]	(0.003)	(0.007)
On Farm	0.21	0.1	0.1	-0.079***	-0.089***
	[0.41]	[0.29]	[0.29]	(0.002)	(0.004)
Urban Area	0.45	0.51	0.52	0.083***	0.069***
	[0.50]	[0.50]	[0.50]	(0.003)	(0.007)
Literate	0.94	0.88	0.86	-0.060***	-0.070***
	[0.23]	[0.33]	[0.34]	(0.002)	(0.005)
In Labor Force	0.96	0.98	0.98	0.025***	0.021***
	[0.20]	[0.13]	[0.14]	(0.001)	(0.002)
Occupation Score	19.82	21.13	21.34	1.059***	1.200***
	[13.37]	[10.55]	[10.51]	(0.075)	(0.152)
High Skill	0.26	0.16	0.17	-0.110***	-0.098***
	[0.44]	[0.36]	[0.37]	(0.003)	(0.005)
Semi Skill	0.19	0.5	0.49	0.284***	0.275***
	[0.39]	[0.50]	[0.50]	(0.004)	(0.007)
Low Skill	0.32	0.25	0.25	-0.077***	-0.071***
	[0.47]	[0.44]	[0.43]	(0.003)	(0.006)
Miner (Occupation)	0.06	0.42	0.41	0.339***	0.326***
	[0.25]	[0.49]	[0.49]	(0.004)	(0.007)
N	7,488,001	19,772	4,804	7,507,773	7,492,805

Notes: The table compares the mean characteristics (and standard deviation in brackets) of fathers residing in Pennsylvania, Illinois, West Virginia, and Ohio during 1900, 1910, and 1920 (Column 1) with fathers linked to mining accident records. All fathers involved in mining accidents and fathers that eventually perish in a mining accident are displayed in Columns 2 and 3, respectively. The right two columns display the differences in means between all fathers and fathers involved in mining accidents, conditional on the census year and state of residence fixed effects. The first difference column displays the differences between all fathers involved in mining accidents and all other fathers, while the second difference column displays the difference between fathers killed in mining accidents and all other fathers. Occupational skill groups follow the 1950 occupation definition of the U.S. Census Bureau. Standard errors are clustered by county of residence and are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Summary Statistics of Sons During Childhood

	Sons of Non-Fatal	Sons of Fatal	Differences
Age at Census	6.06	6.08	0.133
	[4.38]	[4.45]	(0.085)
Age At Accident	10.17	10.18	-0.053
	[4.57]	[4.58]	(0.107)
Young At Accident	0.18	0.18	0.007
	[0.38]	[0.38]	(0.008)
White	0.99	0.99	0.001
	[0.11]	[0.11]	(0.003)
Num. Siblings	3.27	3.40	0.081
	[2.11]	[2.09]	(0.063)
In School	0.50	0.51	-0.010
	[0.50]	[0.50]	(0.009)
Literate	0.22	0.23	0.018*
	[0.42]	[0.42]	(0.007)
In Labor Force	0.02	0.02	-0.000
	[0.14]	[0.14]	(0.003)
N	12,206	4,073	16,279

Notes: The table presents summary statistics for sons of fathers listed in state mining accidents that were linked to the complete count U.S. Census. The first column displays summary statistics for sons of fathers that will eventually be involved in serious, but non-fatal mining accidents. The second column displays the same statistics for sons of fathers that will eventually perish in a mining accident. Finally, the third column presents regression adjusted differences in means between the two groups of sons. Young is defined as the son being younger than primary school age at the time of their father's accident. The specification includes state and census year fixed effects. Heteroskedasticity robust standard errors are presented in parentheses and standard deviations are presented in square brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Estimates of Parental Death By Birth Order

	Birth Order		
	First	Second	Third or higher
Father Fatal Accident	-0.046	-0.102**	0.022
	(0.048)	(0.043)	(0.046)

Notes: The table reports the estimates of parental death on the log of wage income by birth order. The estimates stem from the regression described in Section IV.B, interacting losing a father with birth order and estimating the total effect of losing a father for each group. Standard errors clustered at the adulthood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Heterogeneity by Extended Family Member Income

(1)	(2)
No Available Extended Family	No Available Extended Family
-0.039	-0.042
(0.028)	(0.026)
Extended Family Present	Extended Family with Above Median Income
0.029	0.194*
(0.088)	(0.116)

Notes: The table presents the results of regressing the log of wage income on losing a father across several measures of subgroup heterogeneity. Each regression includes the full set of controls in the preferred specification described in Section IV.B. Each column represents a separate estimation of equation 2 with estimates of γ in the top row and $\gamma+\theta$ below. The first three columns examine differences in the main wage effects by the availability of family support. In Column 1, households are identified by whether or not they lived with extended family. Column 2 identifies households by whether or not they lived with an extended family member who had an occupation-based income score above the national median. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Robustness to Potential Confounders

	(1)	(2)	(3)	(4)	(5)
	Dry Accidents	PA Only	Accident FE	Town FE	Omit Adulthood County
Panel A: Average Effect					
Father Fatal Accident	-0.046*	-0.058**	-0.064**	-0.053**	-0.038
	(0.026)	(0.028)	(0.028)	(0.026)	(0.026)
Panel B: Average Effect, Young					
Fatal × Young	-0.184***	-0.177**	-0.181**	-0.178***	-0.170***
	(0.064)	(0.072)	(0.076)	(0.067)	(0.065)
N	5,526	3,700	3,607	5,325	5,526

Notes: This table reports the results from regressions of the log of wage income in 1940 on losing a father in a mining accident. Each column represents a modified version of the preferred specification described in Section IV.B meant to test the sensitivity of the baseline estimates to potential confounding variation. The first column includes an indicator for whether or not the accident occurred after the enactment of a local prohibition law. The second and third columns examine only accidents that occurred in Pennsylvania, while the third column additionally includes accident-type fixed effects. Instead of childhood county of residence fixed effects, the fourth column includes childhood town of residence fixed effects from Berkes et al. (2023). Finally, the final column omits fixed effects for the son's adulthood county of residence. Standard errors clustered at the adulthood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Robustness to Alternative Inference Methods

	(1) Log(Wage Inc.)	(2) Above Med. Inc.	(3) Inc. Percentile	(4) Non-Wage Inc.
Fatal Accident × Young	-0.172	-0.062	-3.251	0.036
	(0.061)	(0.035)	(1.771)	(0.029)
	p-values from...			
Clustered by Adulthood County	0.005	0.080	0.067	0.213
Clustered by Childhood County	0.005	0.083	0.083	0.094
Two-Way Clustered	0.003	0.083	0.054	0.096
Clustered by Accident Cohort	0.032	0.160	0.140	0.228
Wild-Cluster Bootstrapped Ses	0.001	0.050	0.035	0.228

Notes: This table shows the robustness of the effects of parental death at an early age to alternative inference methods. Each column of the table presents the main estimate, clustered standard errors at the adulthood county level in parentheses, and p-values computed from alternative procedures for each of the measures of income. Alternative p-values are computed using (i) standard errors clustered at the adulthood county level, (ii) standard errors clustered at the childhood county level, (iii) two-way clustered standard errors at both the adulthood and childhood county levels, (iv) standard errors clustered by accident cohort, and (v) wild-cluster bootstrapped standard errors at the adulthood county level.

Table A.8: Estimates of Parental Death on Occscore Measures

	(1) Occscore	(2) CW Occscore	(3) LIDO score
Panel A: Average Effect			
Father Fatal Accident	-0.010	-0.022	-0.001
	(0.011)	(0.017)	(0.010)
Panel B: Average Effect, Young			
Fatal × Young	-0.027	-0.038	-0.018
	(0.023)	(0.046)	(0.018)
N	5,511	4,322	5,187

Notes: This table reports the effects of parental death on three measures of occupation-based income score. In the first column, the dependent variable is the log of the standard occupational income score (occscore). In Column 2, I construct an alternate version of occscore in the spirit of Collins and Wanamaker (2014) that adjusts for state, industry, occupation, and urban-rural differences. Finally, in Column 3, I employ an alternate version of the typical occupational income score developed by Saavedra and Twinam (2020), which is based on occupation, industry, demographics, and geography rather than occupation alone. In Section C.C, I discuss the income score measures in more detail. Panel A re-estimates β from Eq. 1 across each dependent variable, while Panel B does the same for $\gamma + \theta$ from Eq. 2, estimating the effects of losing a father at a young age. Each regression includes the full set of controls in the preferred specification described in Section IV.B. Standard errors clustered at the adulthood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Other Specific Occupations

	(1)	(2)	(3)	(4)	(5)	(6)
		Laborer			Farmer	
	Non-Wage Occ.	Laborer (all)	Laborer (n.e.c.)	Farmer	Farm Labor	On Farm
Panel A: Average Effect						
Father Fatal Accident	-0.010	0.010	0.007	0.001	-0.001	-0.003
	(0.015)	(0.015)	(0.015)	(0.004)	(0.004)	(0.008)
Panel B: Average Effect, Young						
Fatal × Young	0.059	0.009	-0.001	0.022	0.010	-0.012
	(0.045)	(0.032)	(0.032)	(0.015)	(0.011)	(0.018)
Control Group Mean	0.301	0.159	0.155	0.026	0.018	0.068
N	5,525	5,525	5,525	5,525	5,525	5,525

Notes: This table reports the results from regressions of working in a specific occupation on losing a father in a mining accident. In the first column, the dependent variable is whether or not the individual worked in an occupation that earned a large share of income from non-wage sources. To calculate the variable, I take an extract of the 1950 census and calculate the share of income derived from non-wage income by each occupation. I then identify the occupations that receive more than 10 percent of income (the median share of non-wage income across all occupations) from non-wage sources. The second and third columns examine laborer occupations, either all classes of laborers or laborers not elsewhere classified. Finally, columns 4 through 6 examine proxies for farm work. Panel A re-estimates β from Eq. 1 across each dependent variable, while Panel B does the same for $\gamma + \theta$ from Eq. 2, estimating the effects of losing a father at a young age. Each regression includes the full set of controls in the preferred specification described in Section IV.B. Standard errors clustered at the adulthood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Parental Death on Mothers' Outcomes Conditional on Sons' Age

	(1) Widowed	(2) Head of Household	(3) Number of Children	(4) Labor Force
Husband Fatal Accident	0.239***	0.243***	-0.052	0.040**
	(0.032)	(0.030)	(0.126)	(0.016)
Observations	4,124	4,124	4,124	4,124

Notes: This table reports the effects of losing a husband on the outcomes of mothers, among the sample of mothers identified by sons who were younger than 18 in the nearest census following an accident. The estimates of these columns are derived from a specification that includes the fathers' pre-accident characteristics and childhood characteristics described in Section IV.B, as well as fixed effects for the initial census year observed, county of residence during childhood, and for the birth year and accident year cohorts. Standard errors clustered by childhood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Effect of Injuries

The identification strategy relies on comparing the sons of fathers who perished in an accident to those who survived a non-fatal accident. As discussed above in Section II, state inspectors typically reported a non-fatal accident if the victim was unable to work for a sustained period of time (typically 30 days). Some survivors, however, experienced severe injuries that kept them out of work for much longer or led to disability. If a severe, debilitating injury diminished the capacity for parental investment in the victim's household, then their son's future may have been affected as well. Put differently, the counterfactual group may be partially treated due to the prevalence of substantial injuries that affect one's future capacity to work and thrive, and as a result, the effects of losing a father could be underestimated. To explore this concern, I examine the effects of injuries on surviving victims directly and estimate the long-run effects of injuries on the sons of survivors.

Long-Term Effect of Injuries on Surviving Victims

While one may expect injuries to result in the average estimated effect of losing a father to be underestimated, the likely small share of debilitating injuries in the analysis sample suggests this may not be the case. As highlighted in Section II, less serious accidents were nearly 42 times as common as fatal accidents, and a contemporaneous survey of non-fatal accidents found that only ten percent of injuries were serious enough to disable a victim (Adams et al., 1932; Roberts, 1904). To directly examine how injuries affected the later employment outcomes of injured fathers, I follow a subset of fathers listed as being injured in a non-fatal mining accident to the nearest census following their injury. Doing so identifies 5,333 working-age fathers who can be found both in the nearest censuses before and after their accident.[29]

Table B.1 describes the injured fathers before and after their accident in Columns 1 and 2,

[29] Since some fathers may have reached retirement age by the nearest census following their accident, I focus on those who were aged 20-50 in the census before their accident.

respectively. Nearly all (99 percent) of the fathers were in the labor force before their accident, and two out of five worked as miners. Similarly, roughly three-quarters are listed in a semi-skilled or low-skilled occupational group. Column 3 presents before-and-after differences, net of county, census year, accident year, and birth cohort fixed effects. Two characteristics stand out following their injury. First, fathers listed in non-fatal accidents were roughly two percentage points less likely to be in the labor force. In other words, few injured fathers were unable to work due to their accident. Second, injuries resulted in a departure from the mining industry; roughly one in three victims worked in the industry in the nearest census following their accident. However, the exit from mining did not substantially alter later employment outcomes: roughly 68 percent of injured fathers still worked in either a semi-skill or low-skill occupation group, and occupational income scores were no different following an accident. Altogether, following injured fathers and their sons suggests that while a serious, non-fatal mining accident was likely a significant short-run shock to victims and their families, the lack of long-run effects on victims themselves suggests intergenerational injury effects were small.

Testing for an Injury Effect on Sons

Although the previous section suggests that the scope for an intergenerational effect of injuries in this context is small, it is worth estimating whether the sons of injured victims were meaningfully affected. Estimating the effects of injuries on victims' sons requires some valid counterfactual. To test if injuries from mining accidents affected sons' long-run economic well-being, I implement a selection-on-observables design that combines exact matching and inverse propensity weights. While this approach lacks the sort of quasi-random variation afforded by the empirical strategy to estimate the effects of death, it allows me to compare sons of non-fatal accident victims to a counterfactual set of sons identified in the same local geography and census year with similar pre-accident household characteristics.

Specifically, I generate exactly matched groups based on the census year in which a son's father was matched, the town they lived in before the accident, and the industry their father worked in at the time of the census.[30] I then restrict the sample to include only observations within non-degenerate groups—exactly matched groups that contain at least one son whose father was listed in a non-fatal accident. To compute weights, I estimate a probit model to regress whether the father was listed in the non-fatal accident records on the father's characteristics and those of the household. The included characteristics are the census year and county in which the household was identified, the father's age and race, marital status, homeownership status, residence on a farm, literacy, and an exhaustive set of occupational, industrial, and birthplace indicators for the father. Estimating the model yields a predicted probability, $\hat{p}$, of being listed in a non-fatal accident. Finally, I implement inverse propensity weights by weighing counterfactual group units by $\frac{\hat{p}}{1-\hat{p}}$.

Figure B.1 plots the distribution of $\hat{p}$ for fathers injured in a mining accident and for those in the counterfactual group as solid and short-dashed lines, respectively. As one would expect, fathers in the counterfactual group, on average, are less likely to be predicted in a non-fatal accident, though the distributions have broad common support. The long-dashed line demonstrates that the differences between the two distributions disappear after weights are applied. Table B.2 similarly supports the selection-on-observables design. The first two columns display average characteristics between fathers injured in an accident and those in the counterfactual group, respectively. While injured fathers are noticeably different from those in the counterfactual, Column 3 shows the differences become small and insignificant after applying the weights and restricting comparisons within exactly matched groups.

To examine whether serious, but non-fatal mining accidents affected the adulthood outcomes of sons, I link the sons of fathers in the counterfactual group to the 1940 census. I then weigh and group each son by their father's corresponding propensity weight and exactly matched group. Table B.3 presents the results of regressing the adult sons' measures of

[30] Rather than restricting comparisons within the county where a household was identified, I use place-based clusters identified by Berkes et al. (2023) to restrict comparisons to households within the same local cluster. Doing so requires comparisons to be made within fine geographic areas.

economic well-being on whether their father was injured in an accident after applying both inverse propensity weights and exactly matched group fixed effects. Across each measure, there is little evidence that a father's injury had a lasting effect; each estimate is insignificant and not meaningfully different from zero.

Finally, it is useful to specifically examine the sons of injured victims who left the mining occupation directly. Their departure from the mine may reflect the severity of an injury (if the accident left the survivor unable to continue in the occupation) or scarring from the industry. In any case, their departure may be useful in informing the average estimates of the long-run effects of losing a father. If sons of victims who left the industry experience similar declines in economic well-being as those who lost their fathers, this could implicate the loss of occupational continuity between father and son, or the loss of access to industry-specific knowledge and labor networks. Conversely, if the sons of these victims experience no long-term effects, this might indicate the importance of the father's presence in the household and the role of parental capital transmission. Table B.4 examines the long-run, adulthood outcomes of the 454 sons whose fathers left mining following a non-fatal accident.[31] Similar to the sons of accident survivors, sons whose fathers left the mining industry following their injury did not experience significantly different economic outcomes in adulthood compared to the matched counterfactual group.

[31]To specifically identify sons of accident survivors who departed the mining industry, I first identify the set of victims who were observed as no longer working in mining in the nearest census following the accident. Specifically, of the 5,333 accident victims who can be traced to the nearest census following their accident, 965 were identified as both working in mining before an accident and departing the industry by the next census. Matching these fathers to their sons, and then matching their sons forward to the 1940 census, identifies 454 individuals.

Table B.1: Estimates of Injury Effects on Non-Fatal Accident Victims

	Pre Accident	Post Accident	Differences
In Labor Force	0.99	0.97	-0.021***
	[0.10]	[0.17]	(0.003)
Miner (Occupation)	0.41	0.3	-0.107***
	[0.49]	[0.46]	(0.017)
Mining (Industry)	0.41	0.32	-0.089***
	[0.49]	[0.47]	(0.015)
High Skill	0.16	0.2	0.031***
	[0.37]	[0.40]	(0.012)
Semi Skill	0.49	0.4	-0.096***
	[0.50]	[0.49]	(0.015)
Low Skill	0.25	0.28	0.035***
	[0.43]	[0.45]	(0.010)
Log(Occscore)	3.15	3.14	-0.006
	[0.30]	[0.33]	(0.006)
Observations			5,333

Notes: The table displays the mean characteristics (and standard deviation in brackets) of working-age fathers listed as being injured in a mining accident who could be linked to the nearest census following their accident. The first column displays the summary statistics for these fathers in the nearest census prior to their accident, while Column 2 displays the same for the nearest census following their injury. Column 3 displays the before-and-after difference, net of county, census year, accident year, and birth cohort fixed effects. Standard errors are clustered at the county where the individual was first observed. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Balance Among Injured Fathers and a Matched Counterfactual Group

	(1) Non-Accident Fathers	(2) Non-Fatal Accident Fathers	(3) Weighted Differences
Age	42.98	38.44	0.102
	[12.63]	[9.76]	(0.152)
Married	0.95	0.97	0.000
	[0.21]	[0.16]	(0.002)
Number of Children	2.79	3.17	-0.004
	[1.80]	[1.96]	(0.019)
White	0.98	0.98	-0.001
	[0.14]	[0.12]	(0.001)
Foreign Born	0.29	0.47	0.005
	[0.45]	[0.50]	(0.011)
Homeowner	0.45	0.37	-0.001
	[0.50]	[0.48]	(0.006)
On Farm	0.21	0.1	-0.003
	[0.41]	[0.29]	(0.002)
Urban Area	0.45	0.5	-0.016***
	[0.50]	[0.50]	(0.005)
Literate	0.94	0.88	0.001
	[0.23]	[0.32]	(0.005)
In Labor Force	0.96	0.98	0.000
	[0.20]	[0.12]	(0.000)
Occupation Score	19.82	21.07	-0.075
	[13.37]	[10.56]	(0.088)
High Skill	0.26	0.15	-0.001
	[0.44]	[0.36]	(0.004)
Semi Skill	0.19	0.5	0.001
	[0.39]	[0.50]	(0.004)
Low Skill	0.32	0.26	0.001
	[0.47]	[0.44]	(0.003)
Miner (Occupation)	0.06	0.42	-0.001
	[0.25]	[0.49]	(0.004)
N	7,488,001	14,968	5,588,119

Notes: The table displays the mean characteristics (and standard deviation in brackets) of fathers residing in Pennsylvania, Illinois, West Virginia, and Ohio during 1900, 1910, and 1920 (Column 1) as well as fathers linked to serious, non-fatal mining accidents (Column 2). Column 3 displays the differences between the two groups by regressing each characteristic on whether the father was injured in a mining accident and after applying inverse propensity weights and conditional on exactly matched group fixed effects. Standard errors are clustered by exactly matched groups and are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Estimates of Injury Effects on Adulthood Economic Well-Being

	(1) Log(Wage Inc.)	(2) Above Med. Inc.	(3) Inc. Percentile	(4) Non-Wage Inc.	(5) Miner
Father Injured	-0.017	0.005	-0.326	0.001	0.006
	(0.012)	(0.007)	(0.407)	(0.005)	(0.006)
Observations	826,811	826,811	826,811	826,811	826,811

Notes: This table reports estimates of the effects of a father being injured in a non-fatal mining accident on the economic well-being of their adult sons. The sample includes the adult sons of fathers listed in a non-fatal mining accident and a counterfactual group consisting of sons whose fathers were identified in the same census year, lived in the same town, and worked in the same industry as the injured fathers. Each column presents the results of a separate regression of a measure of well-being at the top of the column on whether the father was injured in a mining accident and after applying both inverse propensity weights and conditional on exactly matched group fixed effects. Each specification also controls for individual characteristics of sons observed in the 1940 census and characteristics of their fathers and their childhood county. Individual characteristics of adult sons included whether they were in the labor force birth cohort, industry of employment, and county of residence fixed effects. Included controls for fathers are their literacy, nativity, their industry of employment, and occupational income score. Standard errors are clustered at the adulthood county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Injury Effects Among Sons Of Survivors Who Left Mining

	(1) Log(Wage Inc.)	(2) Above Med. Inc.	(3) Inc. Percentile	(4) Non-Wage Inc.	(5) Miner
Father Injured	-0.008	0.009	0.932	-0.012	-0.001
	(0.045)	(0.023)	(1.315)	(0.017)	(0.024)
Observations	821,906	821,906	821,906	821,906	821,906

Notes: This table reports estimates of the effects of a father's injury in a non-fatal mining accident on the economic well-being of their adult sons, specifically among the sample of sons whose fathers exited the mining industry after the accident. The sample includes the adult sons of fathers listed in a non-fatal mining accident and a counterfactual group consisting of sons whose fathers were identified in the same census year, lived in the same town, and worked in the same industry as the injured fathers. Each column presents the results of a separate regression of a measure of well-being at the top of the column on whether the father was injured in a mining accident and after applying both inverse propensity weights and conditional on exactly matched group fixed effects. Each specification also controls for individual characteristics of sons observed in the 1940 census and characteristics of their fathers and their childhood county. Individual characteristics of adult sons included whether they were in the labor force birth cohort, industry of employment, and county of residence fixed effects. Included controls for fathers are their literacy, nativity, their industry of employment, and occupational income score. Standard errors are clustered at the adulthood county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

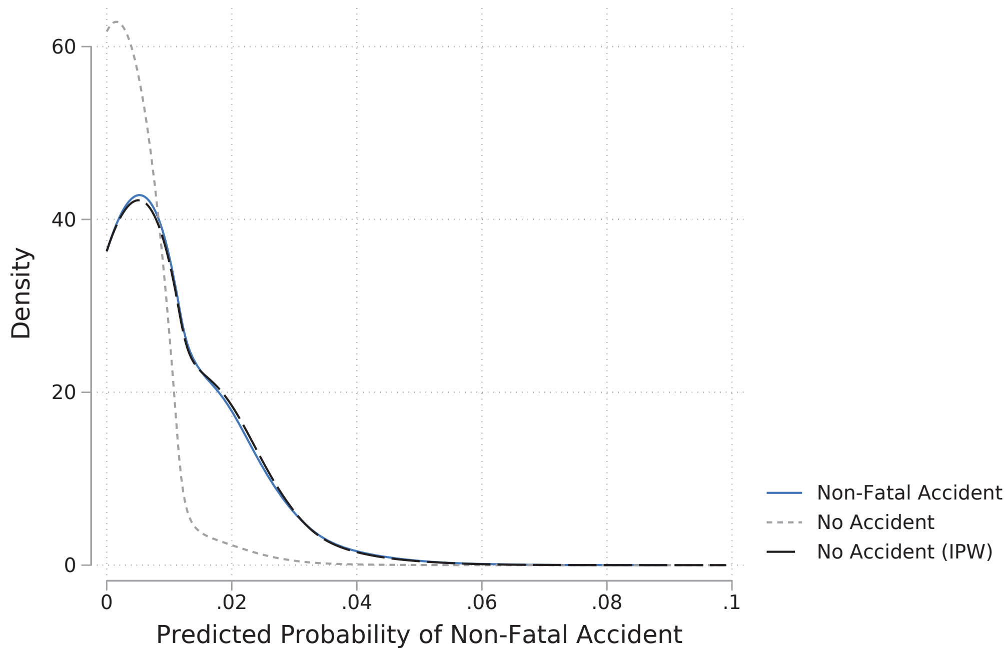


Figure B.1: Distribution of predicted propensity for non-fatal accident involvement

Notes: The figure displays the distribution of predicted propensity for having a father injured in a mining accident. The predicted propensities are estimated via a probit model, and the full set of predictors included are race, birth year, whether the father was literate, the number of children identified in the household, the census year and county the father was identified prior to the accident, and an exhaustive set of dummies for the father's occupation, industry, and place of birth.

C Data Appendix

This section provides additional details on the construction of the main datasets and variables employed in the analysis. Each section discusses the various sources for the data and the methods used to compile the raw files into their final forms.

C.A Accident Records

Digitized mining accident records come from two primary sources: The Gerald E. Sherard collection of the Colorado School of Mines and collection RG-45 of the Pennsylvania State Archives. Each collection compiles individual mining accidents across several sources, including state mine inspection records, contemporaneous local news articles, websites dedicated to the history of mining in several local communities, and the descendants of the victims. The quality and completeness of the individual accident records vary across sources. To build the dataset, I retain only records that are reported in official state inspection reports from Pennsylvania, Illinois, West Virginia, and Ohio. Altogether, I collect 217,457 individual mining accident victims from these four states.

Individuals may appear in the records a number of times if they experienced multiple non-fatal accidents throughout their careers. Additionally, a common challenge with historical registers is that it is impossible to separately identify, since the data contain no unique administrative identifiers and individuals with common names that share a birth date. For example, Lahey Coleman born in 1900 may appear in an accident in 1922 and 1923 or even multiple times in the same year. However, historical record linking requires unique individuals, typically by name, birth year, and some level of geography (e.g., state of birth or residence).

To create a dataset compliant with the linking techniques utilized in this paper, I assemble a dataset that is unique by state, full name, and birth year. Table C.1 describes the step-by-step criteria for dealing with duplicates and creating records compliant with census linking methods. Starting with the original set of 217,457 records, I exclude any accident where the age is missing or invalid (e.g., includes non-numeric characters). I also exclude records where the accident date or full name is not complete (e.g., no last name). Since the records are compiled from several inspectors, some inspectors recorded the same victims on the same dates, I also drop duplicates in this case.

From the remaining set of records, I keep the terminal fatal accident or the latest non-fatal accident. However, this step is complicated by the fact that names (and birth years) may not be unique. In other words, the data may contain an individual who is listed in a fatal accident in 1915, and a non-fatal accident in 1916. This is particularly true for individuals with common names. As a result, if there is an individual with a fatal accident prior to a non-fatal accident at any point, they are not included in the sample.

Finally, I retain only records that correspond to accidents that occur between 1900-1929. Doing so results in a dataset of 183,536 individuals unique by name, birth year, and state in which the accident occurred. Each record corresponds to one individual and their final accident observed in the state records.

Accident Causes

The nature and cause of accidents provide a brief (and occasionally detailed) description of the incident that led to fatality or a serious, but non-fatal injury. The digitized descriptions of accidents are only available for the 110,699 records from Pennsylvania. Given the complex descriptions of some of the accidents, I group causes into five main categories: Fall of Rock, Cars and Machines, Explosions and Gas, Falls of Persons or Tools, and Other.

The first category consists of all accidents from falling rocks, coal roofs, or spoil tips. The most frequent descriptions in this category are simply "fall of rock," "fall of slate," or "fall of roof." I use a set of keywords such as *rock, slate, roof, fall of coal, collapse, rush of culm,* or *smothered by rock* to identify accidents where sudden falls of rock are the primary cause of an accident. Accidents involving collapses of roofs or other falling rocks account for nearly half of all mining accidents.

Accidents caused by Cars and Machinery contain all incidents involving mine cars, steam engines, rail cars, and various machines employed in processing coal. Such accidents are typically described simply as "by cars,""crushed by cars," "run over by cars," or involve being caught between machinery. I identify accidents caused by cars and machinery by relying on keywords such as *cars, motors, machines, breaker, engine, railcar, run over, off track,* or *faulty brake*. Over one quarter of all mining accidents are caused by cars or machinery and the majority of those accidents are the fault of cars.

Explosions, faulty blasting shots, and gas account for roughly thirteen percent of all accidents. These accidents are more likely to result in fatality and are responsible for a number of mass casualty events throughout the early 20th century. To identify accidents caused by explosions or gas I search for keywords like *gas, charge, explosion, blast, powder, shot* or *dynamite*. These accidents are typically described as premature or delayed blasts, flying rocks from explosions, gas ignited by fire or arcs of electricity, or collapsed coal rooms from explosions.

The fourth category contains all accidents that list falling persons or tools as the primary cause of accident. These accidents mostly include individuals slipping and falling while in the mine or fault a fall in the mine cage, a conveyance used to transport workers and supplies from the surface to the mine by way of a rope. Keywords used to identify falls include *fall of person, slipped, fell off, fell down, fell under,* or *fall of cage*. These accidents account for eight percent of all mining accidents.

Finally, accidents not elsewhere classified are included in the Other category. A catch-all, this category consists mostly of accidents cause by animals, suffocation, drowning, or electrocution and accounts for the remaining set of accidents.

Identifying accident causes using keywords is an imperfect measure of assigning the cause of accident since individual accident causes may be identified as having multiple sources. In some cases, I can reasonably assign fault to cases with multiple sources. For example, "fall of rock from explosion" can safely be assigned to the category of explosions and not the sudden fall of rock. However, some 2,000 incidents have multiple causes that cannot be isolated as the primary cause. For example, "by car and rock" has no clear primary cause. In these few cases, I allow the incident to have multiple causes.

C.B Linking Process

I employ commonly used techniques to historical record linking. Specifically, I employ a refined version of the Ferrie (1996) algorithm developed by Abramitzky et al. (2021). To begin, I resolve discrepancies in naming conventions between census data and the accident records by systematically standardizing first names and replacing commonly abbreviated names with full names. For example, nicknames and shorthand are common in the accident records (E.g. JNO instead of John or WLM instead of William).[32] I define candidate links based on age, state of residence, and similarity scores computed by first and last name. Similarity scores are Jaro-Winkler string distances which give a measure of the similarity between two strings, placing more weight on similar characters at the beginning of the string. A similarity score of 0 represents a candidate whose name exactly matches, while a score of 1 represents two names with no shared characters.

The algorithm relies on fixed information, typically place of birth, to refine the pool of potential links between sources. Since the accident records do not report place of birth, I instead (i) assume individuals listed in a given state's accident records lived in that state at the time of the census and (ii) search for individuals in the nearest prior census. For example, if an individual accident occurred in Pennsylvania during 1904, I search for that accident victim among Pennsylvania residents in the 1900 U.S. Census.

Similar to the procedure outlined in Abramitzky et al. (2021), I implement the following algorithm:

[32] I do not rely on phonetic name cleaning algorithms like Soundex or NYSIIS since Bailey et al. (2020) show that use of phonetic algorithms may increase false positive links.

1. For each observation in dataset A, I define a set of candidate matches in dataset B. An observation in dataset B is a candidate match for an observation in A if:

    - It has the same state of residence.
    - The birth year is within plus/minus two years of the reported birth year in dataset A.
    - The initials of first and last names are the same between candidates.

2. For each pair of candidate matches, I compute the Jaro-Winkler score for the first and last name. I consider a candidate pair matched on names if the total Jaro-Winkler score is less than or equal to 0.10. After computing name similarity scores there are three possibilities:

    - There is no matched candidate for a given observation in dataset A. In other words, there are no records in dataset B that have sufficiently similar (Jaro-Winkler score ≤ 0.10) names in the same state and within the age window.
    - The observation in dataset A has a unique matched candidate in dataset B. In this case, the observations are considered linked and are included in the next phase of the analysis.
    - The observation in dataset A has more than one record that is considered matched on names and within plus/minus two years of birth. In this case, I take the candidate with the smallest difference in reported age as the linked record and include it in the next phase of the analysis. If there are still more than one records with the smallest difference in reported age, there is no unique match and the records are not used.

3. Next, I repeat the above steps, matching observations from dataset B to A (instead of from A to B). The sample that continues on to the next phase of the analysis is the

intersection between observations that match uniquely from dataset A to dataset B and those that match uniquely from dataset B to dataset A.

Linking the accident records to the complete count census years provides a total of 53,510 matches from the 183,536 accident records. The match rate of 29.1 percent is within the standard matching rate in the historical census linking literature which frequently finds match rates between 20-30 percent. Of the 53,510 accident victims matched to the census, I identify 19,772 victims who were listed as fathers at the time of enumeration. From these heads of household, I collect 16,279 native-born sons to link forward to their adulthood outcomes in 1940. I focus only on native-born sons for two reasons. First, I focus on sons since linking daughters presents the additional challenge of marriage. Since women customarily change their surname after marriage, it is not possible to link daughters unless one possesses information about renaming post-marriage. Second, I focus only on native-born children since nearly 98 percent of those identified as children of accident victims are born in their father's state of residence.

In the next step, I link sons of mining accident victims forward to the 1940 Census. For the 16,279 native born sons, I repeat the record-linking algorithm described above; I link each group of sons found in their respective census year forward to the 1940 Census. I then focus only on sons who were younger than 18 at the time of their father's accident since sons who reach young adulthood are less likely to be co-residents during the accident. Moreover, children who were older than 18 and still co-resident with their parents may differ in unobserved ways, and these differences may reflect some sort of underlying qualities that matter for long-run outcomes. Focusing on adult sons who were younger than 18 at the time of their father's accident yields 5,624 men observed in 1940 with valid responses in all outcomes and control variables. Of these adult children of accident victims, 23.8 percent lost their father due to a mining accident.

Selection into matching presents a potential bias if records that match fail to capture the characteristics of the average miner involved in an accident. To probe this potential bias, I compare records that match to the 1900, 1910, and 1920 complete count census with records that fail to find a match in any census year. Table C.2 presents the mean characteristics of matched and not-matched accident records as well as differences between the two means. Accident victims that successfully matched to the complete count census were younger and less likely to perish as a result of their mining accident. Figures C.1a and C.1b plot the distribution of age at the time of accident and birth year of linked and not linked accident victims. These figures provide further evidence that matched and not-matched victims are not substantially different. Finally, Figure C.2 displays the differences between matched and not-matched accident victims standardized by the mean and standard deviation of the not-matched victims in gray. The figure highlights the magnitudes of the differences; differences are clustered around less than ten percent of a standard deviation and are precisely estimated.

Still, to account for potential biases due to record linking, I compute inverse propensity weights (IPWs) for each individual listed in the accident records. To do so, I first estimate a probit model where the dependent variable indicates whether the individual was matched to a census record. In this model, I predict the probability of linking with the string length of the victim's first and last names, whether or not they perished in an accident, and accident year, accident month, state of residence, and birth year fixed effects. I then predict the conditional probability of being linked and construct IPWs as suggested by Bailey et al. (2020). Weighted differences are shown in green in Figure C.2 and are smaller in magnitude than unweighted differences. Finally, Table C.3 compares the main estimates of the paper to estimates weighted by the computed IPWs. The weighted and unweighted estimates are similar in both magnitude and precision, suggesting that selection into the linked sample poses no significant threat to the main identification strategy.

C.C Complete Count Census Files

I rely on full-count data files for the 1900, 1910, 1920, and 1940 U.S. Census. The Census files are provided by IPUMS-USA via a restricted-use licensing agreement. Importantly, the restricted access files contain full names of individuals in the U.S. Census. In this section, I describe the variables employed in the main analyses and their construction.

Income Measures

In order to construct the log of annual wage income reported in 1940, I rely on an IPUMS Census variable, *incwage*, which reports each respondent's total pre-tax wage and salary income for the previous year.[33] Sources of income include wages, salaries, commissions, cash bonuses, tips, money received from public emergency work, and other money income received from an employer. Importantly, this variable measures only money received as an employee during the previous year and does not capture income from assets or other factors of total compensation. In the main analysis, I exclude any respondent who did not report positive annual wage income and re-code top-coded values as $5,000.

While the 1940 Census was the first to report wage income, enumerators were not instructed to measure income from other sources. Instead, respondents were simply asked to report if they received at least $50 of non-wage income. In the 1940 census, non-wage income was meant to capture business profits, professional fees, rent, interest, dividends, pensions, annuities, royalties, and all payments-in-kind. Non-wage income also included unemployment compensation, as well as government and private relief. I use this response to create a dummy variable for whether or not the respondent reported earning at least $50 of non-wage income.

Occupation-based Income Score

Before the 1940 census, data on individual wage income was unavailable in the U.S. Census. As a result, researchers seeking to understand labor market outcomes of individuals relied on occupation-based indices of economic outcomes. Arguably the most popular measure is the occupational-based income score (occ. score), which measures the median income of an

[33] For further details, see `https://usa.ipums.org/usa-action/variables/INCWAGE#codes_section`

occupation in 1950.

While the classic occ. score provides a reasonable proxy for occupational standing rather than individual earnings, the measure may miss important heterogeneity due to geographic and demographic characteristics.[34] Instead of the typical occ. score, I also employ alternative measures suggested by Collins and Wanamaker (2014) and Saavedra and Twinam (2020).

To construct occupation scores in the spirit of Collins and Wanamaker (2014). I retrieve a 1% sample of the 1950 census and examine the average total income of working-age men (aged 20-60) within cells of state, urban/rural status, industry and occupation, and employment status (I do not stratify by race, since 99 percent of the accident sample is White). I then take the log of this adjusted income score and match it to the analysis sample by the specified cells. I also employ a LASSO-adjusted measure of occupation score generated by Saavedra and Twinam (2020). They implement a machine learning approach to construct a new income score based on industry, occupation, geography, and demographics.

Occupation Codes

In 1940, Census enumerators were asked to record respondents' occupations and industry of employment. The responses generated tens of thousands of occupations and industries with limited guidance for comparability. IPUMS has aggregated most of these ambiguous occupations into roughly 200 granular and nine broad categories.[35] The broad categories, along with some example occupations from each class, are:

1. Professionals and Technical Workers: professors, engineers, and healthcare providers.

2. Farmers: farm managers, owners, and tenants.

3. Managers, Officials, and Proprietors: store managers, public administration inspectors, building managers.

[34] For instance, Bailey and Collins (2006); Collins and Wanamaker (2014); Inwood et al. (2019) and Saavedra and Twinam (2020) all highlight that occupational income measures that incorporate geographic and demographic characteristics may better reflect the income standing of individuals within occupations.

[35] For details, see `https://usa.ipums.org/usa-action/variables/OCC1950`

4. Clerical and Kindred: clerks, bookkeepers, office machine operators.
5. Sales Workers: insurance agents and brokers, and salesmen.
6. Craftsmen: blacksmiths, carpenters, and tailors.
7. Operatives: mine operatives and laborers, sawyers, and truck drivers.
8. Service Workers: housekeepers, bartenders, and wait staff.
9. Laborers: farm laborers, lumbermen, day laborers.

From these nine broad occupation categories, I further group occupations into high- semi- and low-skilled occupations. High-skilled occupations are comprised of professionals, managers, and craftsmen. Semi-skilled occupations are comprised of sales workers, clerical workers and operatives. Low-skilled occupations include farmers, service workers, and laborers.

Employment

In Section V, I rely on a series of variables in order to examine the effects of early parental loss on employment outcomes both along the extensive and intensive margins. These variables measure whether respondents were part of the labor force, unemployed, worked a public relief job, and the number of weeks employed during the previous year. This section briefly describes the construction of these variables and the source variables in the 1940 Census records made available through IPUMS (Ruggles et al., 2021).

The primary variable used to construct measures of employment is the respondent's employment status. This categorical variable indicates whether the respondent was part of the labor force (i.e. working or unemployed and seeking work) and, if so, whether the respondent was currently at work. From this categorical variable, I construct simple indicators that mark if a respondent was part of the labor force and if a respondent was currently unemployed.

Importantly, Census enumerators in 1940 were instructed to ascertain whether respondents were either at work on, or assigned to, public emergency work projects conducted by the Work Progress Administration, National Youth Administration, Civilian Conservation Corps, or through some state or local work relief agency. Such respondents were classified as being "on public emergency work." For these respondents, I create an indicator variable that identifies workers who reported their employment status as being affiliated with a public emergency work project.

It is worth noting the degree of misclassification and under-counting surrounding emergency work in the 1940 Census. For instance, the numbers of public emergency workers reported in the United States was just over 2.5 million, while the number recorded by the combined payrolls of the Federal agencies dedicated to public work was over 3.3 million (Census Bureau, 1946). The Census Bureau noted confusion on the part of both enumerators and respondents regarding the classification of certain types of emergency work, and the reluctance of some to report their status as on public work relief. Indeed, the most common type of under-counting occurred when those workers on emergency work listed themselves as "at work" rather than working on a public emergency project.

Finally, enumerators asked respondents to count all weeks over the last year in which any work was done. Enumerators were instructed to calculate the number of full-time equivalent weeks the respondent worked for profit, pay, or as an unpaid family worker during the previous year. For workers, total weeks at work included paid vacations and other paid absences.

C.D Placebo Estimates

Statistical inference is further complicated in this setting since models with relatively few treated units and clusters can lead to improper inference (Cameron et al., 2008; Ferman and Pinto, 2019; MacKinnon and Webb, 2017). It is common to assume that the error term is correlated within clusters, but uncorrelated between them, yet test statistics based on a cluster-robust variance estimator tend to over-reject the null hypothesis when the number of clusters is small. While the wild cluster bootstrap estimator proposed by Cameron et al. (2008) can often lead to more reliable inference, MacKinnon and Webb (2017) show that it can still lead to improper inference if the number of treated clusters is relatively small. In this particular setting, one may share this concern. For instance, while constructing standard errors at the adulthood county of residence level yields 594 clusters, the relatively few sons who experienced the death of a father during early childhood are represented in just 114 of the clusters.

To address this concern, I follow Chetty et al. (2009) and conduct a non-parametric placebo test of the effect of early parental loss on adulthood wage income. To do so, I link nearly 1.6 million sons born in Pennsylvania, West Virginia, and Ohio whose fathers were not involved in a mining accident. Of these non-accident sons, I randomly draw a sample of equal size to the actual analysis sample above and assign placebo indicators to mimic treatment group assignment. Specifically, I randomly assign a placebo indicator for parental loss and being young at the time of an accident to 25 percent and 18 percent of these sons, respectively.[36] Following randomization, I estimate variants of Eq. 2 where the $FatherFatal_i$ and Z_i indicator variables are replaced with the alternative, placebo indicator variables. I repeat this exercise 1,000 times, which yields a distribution of $\widehat{\gamma + \theta}$. The distribution of placebo estimates represents the sampling distribution of $\widehat{\gamma + \theta}$. To compute the p-value associated with the null hypothesis that the estimated income difference among those who lost their parent during early childhood is no different from the placebo estimates, I calculate the percentile of the actual estimate in the distribution of placebo estimates. Since this exercise makes no parametric assumptions about the variance-covariance matrix, nor does it suffer from biases arising from small numbers of treated clusters, I view this as an alternative and conservative approach to statistical inference.

[36] In Table A.1, I show that roughly one-quarter of victims listed in the mine accident records perished as a result of their accident. Additionally, Table A.3 shows that 18 percent of children were less than primary school age at the time of their father's accident.

The first panel of Figure C.3 shows the empirical cumulative distribution functions of placebo estimates (as gray dots), as well as the actual estimates (as a blue diamond). The figure highlights that the actual estimate of parental loss on adulthood income is much larger in absolute value than the placebo estimates and remains statistically significant at conventional levels. The remaining panels of the figure similarly show the empirical cumulative distribution function of placebo estimates for the other measures of income.

Table C.1: Criteria for Analysis of Accident Records

Selection Criterion	Retained Records
Original set	217,457
Drop if age listed in accident records is missing	203,223
Drop if missing full accident date or full name	200,080
Drop duplicated records within the same date	192,556
If fatal comes before non fatal, drop all. If fatal is the final record, keep the fatal accident. Otherwise, take the latest accident.	191,231
Keep accidents between 1900-1929	183,536

Notes: The table describes the criteria used to assemble the unique accident records from the full set of state mine inspectors' reports. The original set of data includes fatal and non-fatal mining accidents from Pennsylvania, Illinois, West Virginia, and Ohio. To create a set of records compliant with historical record linking, the data is transformed into a set of individuals unique by state of accident, full name, and year of birth.

Table C.2: Comparing Matched and Not Matched Records

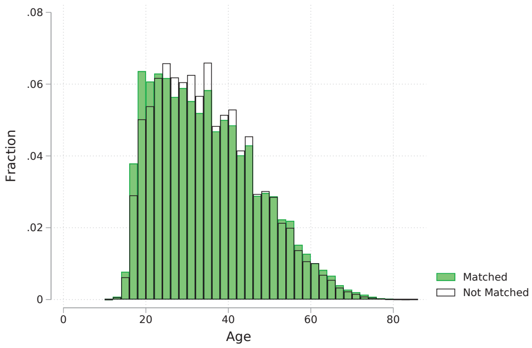
	All	Matched	Not Matched	Difference
Age	34.46	34.36	34.5	-0.134**
	[12.10]	[12.51]	[11.93]	(0.063)
Birth Year	1880.79	1881.69	1880.42	1.277***
	[13.12]	[13.68]	[12.86]	(0.069)
Accident Year	1915.24	1916.05	1914.91	1.143***
	[7.87]	[7.91]	[7.83]	(0.041)
Fatal Accident	0.26	0.24	0.26	-0.024***
	[0.44]	[0.43]	[0.44]	(0.002)
Pennsylvania	0.64	0.66	0.63	0.031***
	[0.48]	[0.47]	[0.48]	(0.002)
Illinois	0.23	0.23	0.23	-0.001
	[0.42]	[0.42]	[0.42]	(0.002)
West Virginia	0.11	0.08	0.12	-0.033***
	[0.31]	[0.27]	[0.32]	(0.001)
Ohio	0.03	0.03	0.03	0.003***
	[0.17]	[0.18]	[0.17]	(0.001)
Fall of Rock	0.45	0.43	0.46	-0.028***
	[0.50]	[0.50]	[0.50]	(0.003)
Cars or Machinery	0.29	0.31	0.28	0.033***
	[0.45]	[0.46]	[0.45]	(0.003)
Explosion or Gas	0.13	0.11	0.13	-0.019***
	[0.33]	[0.32]	[0.34]	(0.002)
Fall of Person or Tools	0.08	0.09	0.08	0.007***
	[0.27]	[0.28]	[0.27]	(0.002)
Other Causes	0.05	0.05	0.05	0.006***
	[0.22]	[0.23]	[0.21]	(0.001)
N	183,536	53,510	130,026	183,536

Notes: The table shows summary statistics for accident records that successfully linked to a record in the complete count census in 1900, 1910, and 1920 and for records that fail to link. Columns 1, 2, and 3 each present the mean and standard deviation in square brackets) of several variables for all accident records, accident records linked to the census, and for those records that fail to link, respectively. The right-most column displays a difference in means between not matched and matched records, along with heteroskedasticity robust standard errors in parentheses. Digitized accident records come from the Gerald E. Sherard Collection at the Colorado School of Mines and are collected from 1900 to 1929. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

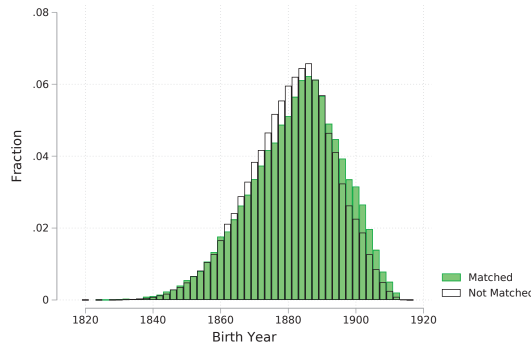
Table C.3: Weighted Estimates of Parental Death on Income Measures

	(1)	(2)	(3)	(4)
	Log(Wage Inc.)	Above Med. Inc.	Inc. Percentile	Non-Wage Inc.
Panel A: Average Effect				
Father Fatal Accident	-0.052*	-0.018	-0.839	0.037***
	(0.028)	(0.015)	(0.974)	(0.013)
Panel B: Average Effect, Young				
Fatal×Young	-0.197***	-0.077**	-3.979**	0.043
	(0.068)	(0.036)	(1.918)	(0.031)

Notes: This table reports the results from regressions of parental death on measures of income in 1940, weighted by propensity weights generated to account for selection into the linked sample discussed in Section C.B. Panel A re-estimates β from Eq. 1 across each dependent variable, while Panel B does the same for $\gamma + \theta$ from Eq. 2, estimating the effects of losing a father at a young age. Standard errors clustered at the adulthood county are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



(a) Histogram of Age by Matched



(b) Histogram of Birth Year by Matched

Figure C.1: Histograms (Matched Vs. Not Matched)

Notes: Figure presents the distributions of age at the time of accident and birth year within the individual accident records by whether or not the record matched to the 1900, 1910, or 1920 Census. The distributions of matched accident victims are presented by green bars while the distributions of non-matched accident victims are presented by hollow bars. Individual accident records come from the Gerald E. Sherard Collection at the Colorado School of Mines.

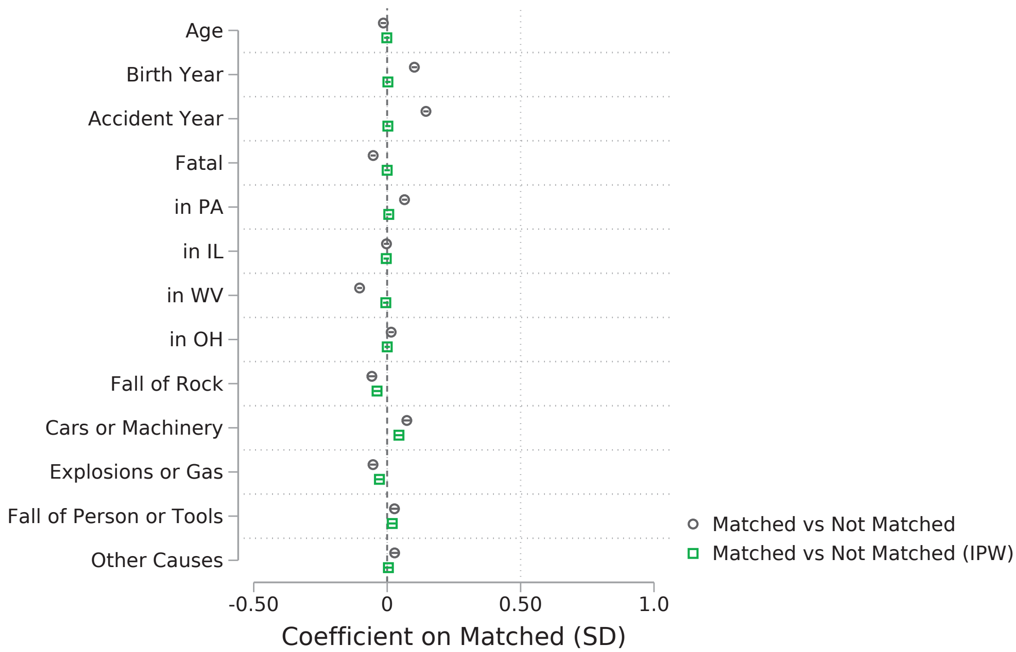
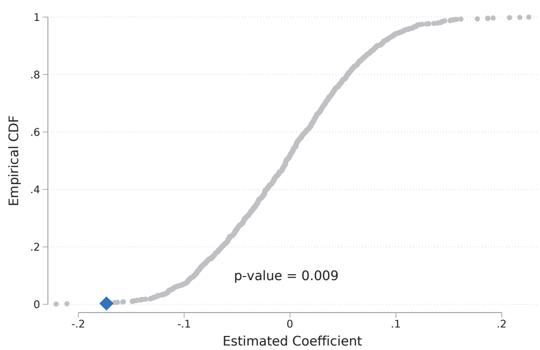
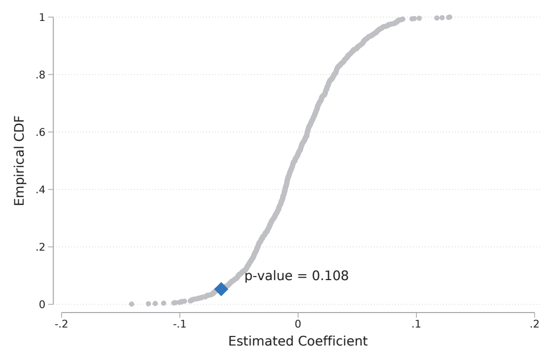


Figure C.2: Difference in Means, SD of Not Matched

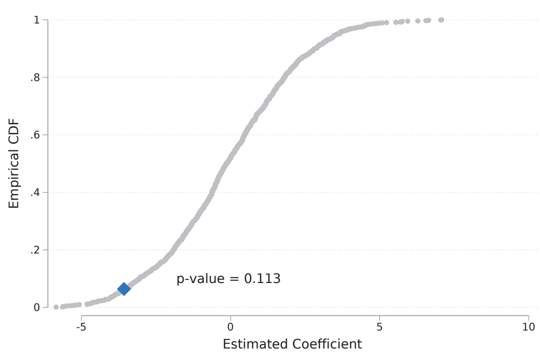
Notes:The figure shows comparisons between matched and not matched records. I standardize variables using the mean and standard deviation of the not matched group. I plot the standardized coefficient (along with 95 percent confidence intervals) from a regression of each characteristic on a dummy for accident victims that match to a census record in 1900, 1910, or 1920 in gray. In green, I plot standardized differences weighted by inverse propensity weights suggested by Bailey et al. (2020). Individual accident records come from the Gerald E. Sherard Collection at the Colorado School of Mines.



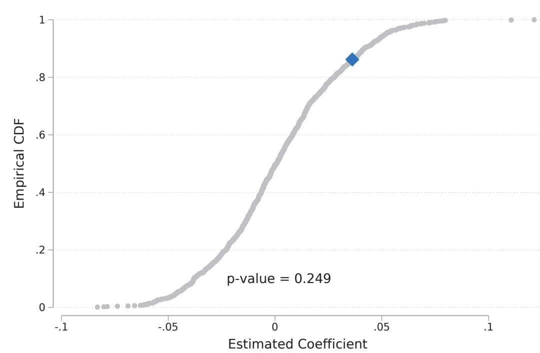
(a) Log(Wage Income)



(b) Above Median Income



(c) Income Percentile



(d) Non-Wage Income

Figure C.3: Distributions of Placebo Estimates

Notes: The figure presents the distributions of placebo estimates. Panels A through D show the empirical cumulative distribution functions of placebo estimates (as gray dots), as well as the actual estimated $\gamma + \theta$ from Eq. 2 (as a blue diamond), separately for each measure of income. Placebo estimates are obtained by estimating 1,000 variants of Eq. 2 and are discussed in more detail in Section C.D.