

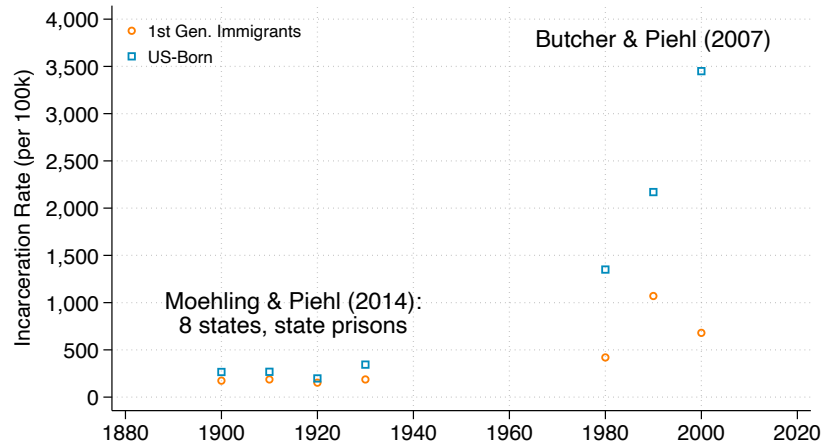
Appendix (for online publication)

LAW-ABIDING IMMIGRANTS: THE INCARCERATION GAP BETWEEN IMMIGRANTS AND THE US-BORN,
1870–2020

Authors: Ran Abramitzky, Leah Boustan, Elisa Jácome, Santiago Pérez, and Juan David Torres

Online Appendix A: Appendix Figures and Tables

Figure A1: Existing Evidence on Immigrant and US-born Incarceration Rates



Notes: This figure plots historical incarceration rates of immigrants and US-born individuals from Moehling and Piehl (2014) as well as modern incarceration rates from Butcher and Piehl (2007). The historical incarceration rates are based on US-born and immigrant individuals ages 18-44 who were incarcerated in state correctional facilities in eight states: Massachusetts, Connecticut, New York, New Jersey, Pennsylvania, Illinois, Michigan, and California. The modern incarceration rates correspond to institutionalization rates among all US-born and immigrant men ages 18-40 from sub-samples of the decennial Censuses.

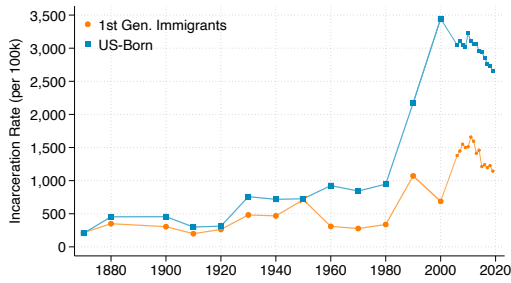
Figure A2: Example Record of Incarcerated Individuals in 1930 Census

E m- of ily er is- ion	<p align="center">NAME</p> <p align="center">of each person whose <i>place of abode</i> on April 1, 1930, was in this family</p> <p align="center">Enter surname first, then the given name and middle initial, if any</p> <p align="center">Include every person living on April 1, 1930. Omit children born since April 1, 1930</p>	<p align="center">RELATION</p> <p align="center">Relationship of this person to the head of the family</p>	<p align="center">Home owned or rented</p>
	5	6	7
	Hardy Frank W	inmate	
	Barrow Clyde	inmate	
	Bowley Pat	inmate	
	Williams Travis	inmate	
	Cornett William L	inmate	

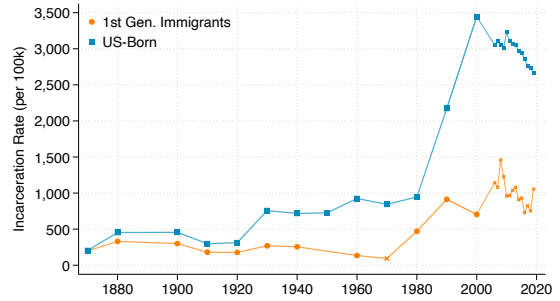
Notes: This figure shows an example record of incarcerated individuals in the 1930 population Census.

Figure A3: Incarceration Rates of Immigrants and US-born Men for 1870-2019, Including 2000

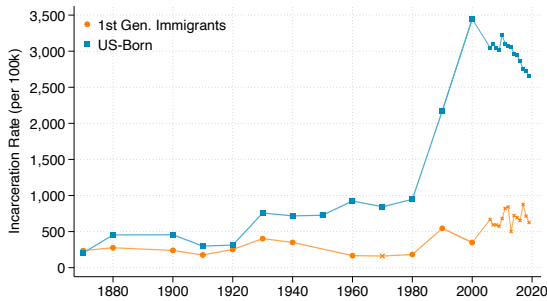
(a) All Immigrants



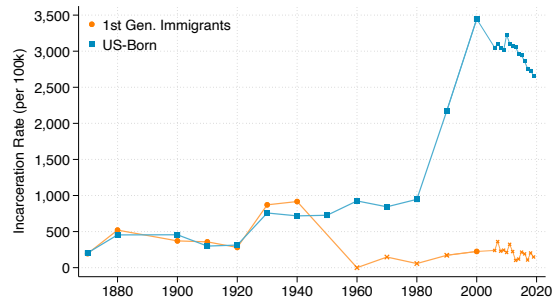
(b) “Old” Europeans



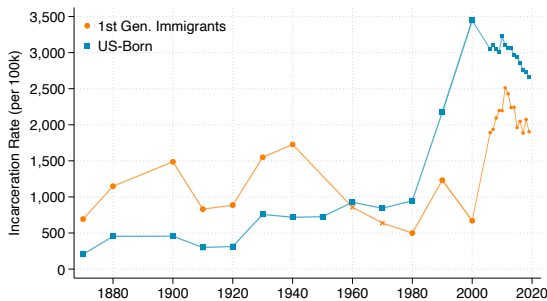
(c) “New” Europeans



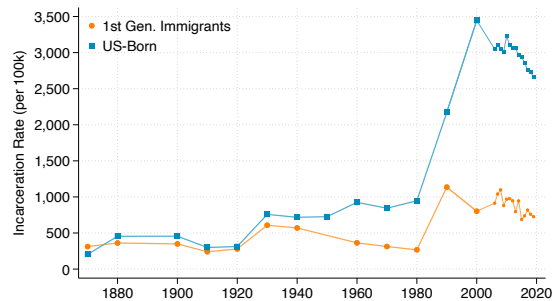
(d) Chinese



(e) Mexican and Central Americans



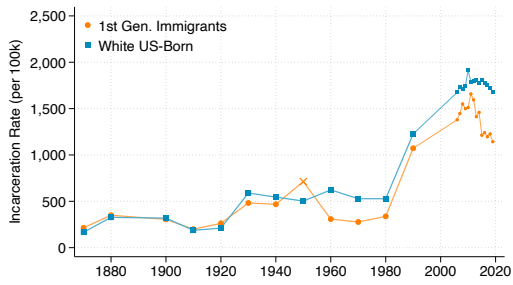
(f) Rest of the world



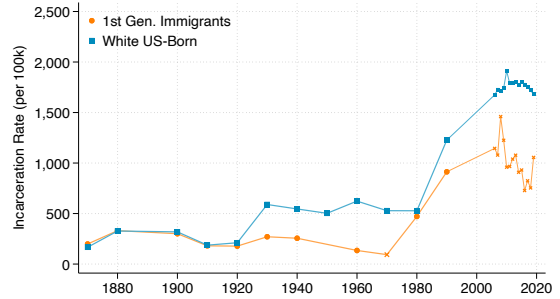
Notes: Each of the panels in this figure plots incarceration rates for immigrants and US-born between 1870 and 2019 as in Figure 1, but including the corresponding points for the 2000 Census. For more details, see the note to Figure 1 and Online Appendix B.

Figure A4: Incarceration Rates of Immigrants and White US-born Men, 1870-2019

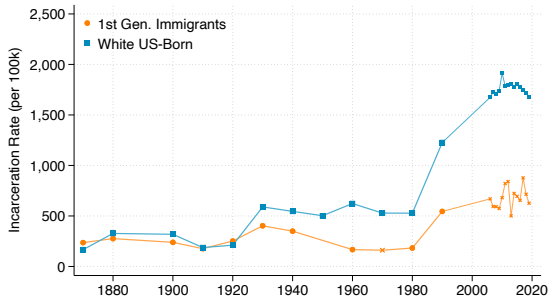
(a) All Immigrants



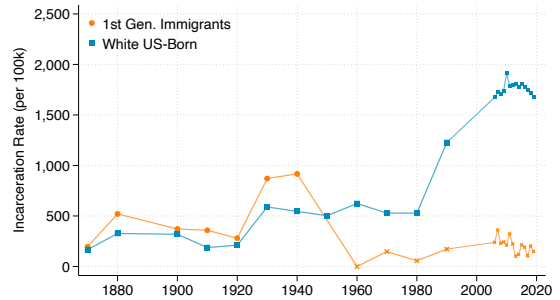
(b) “Old” Europeans



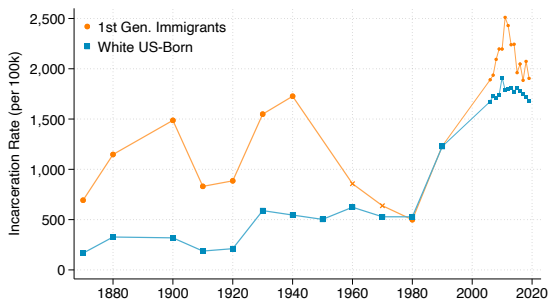
(c) “New” Europeans



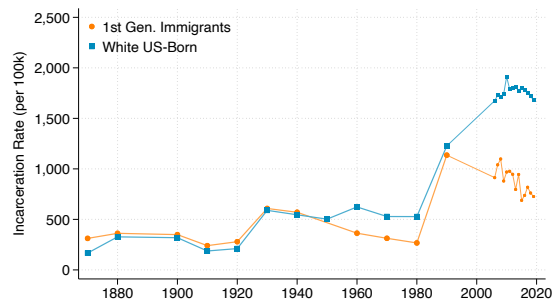
(d) Chinese



(e) Mexican and Central Americans



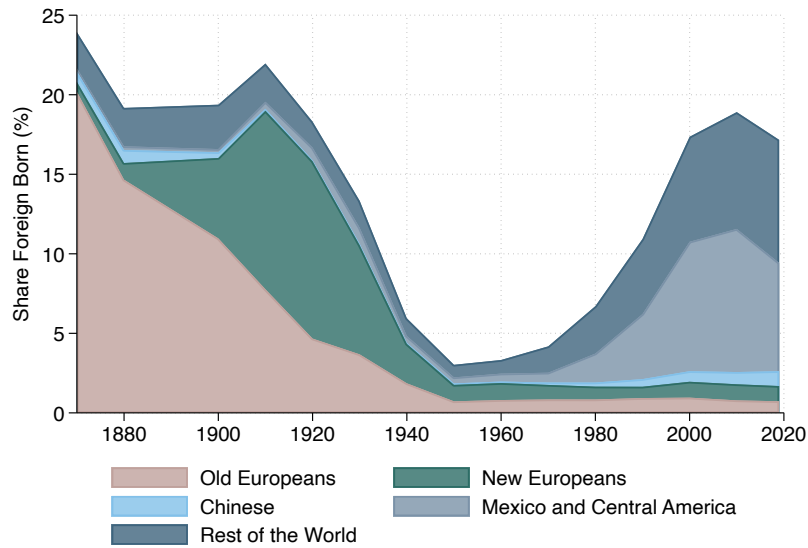
(f) Rest of the world



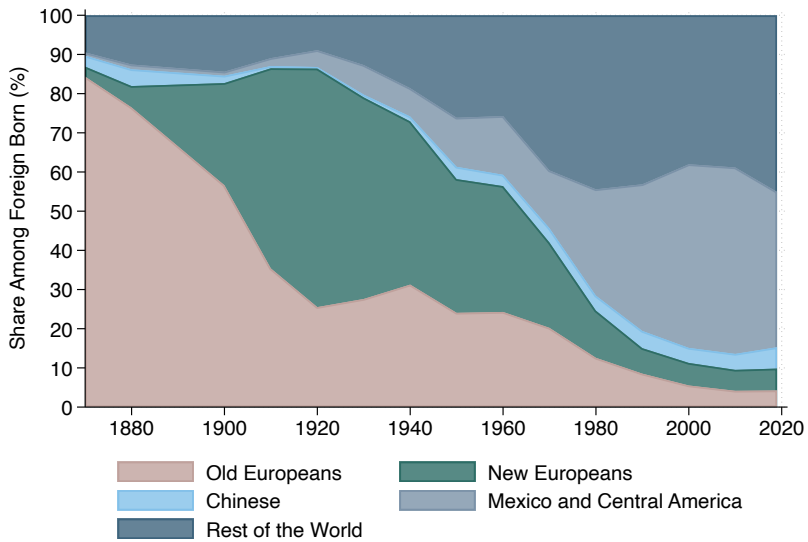
Notes: Each of the panels in this figure plots incarceration rates for immigrants (regardless of their race) and white US-born men between 1870 and 2019. For more details, see the note to Figure 1 and Online Appendix B.

Figure A5: Immigrant Composition in the US, 1870-2019

(a) Within the US Population

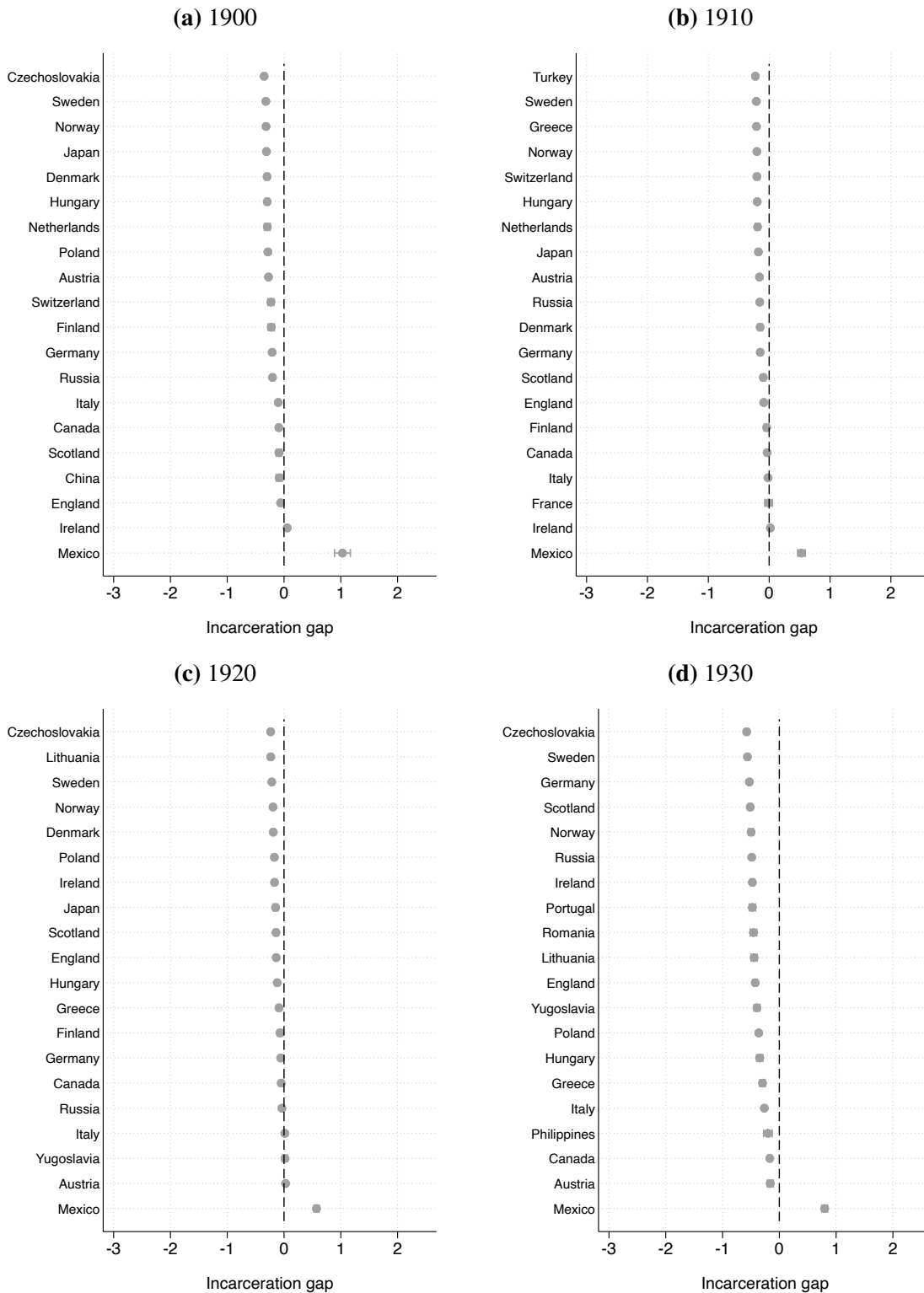


(b) Within the Immigrant Population



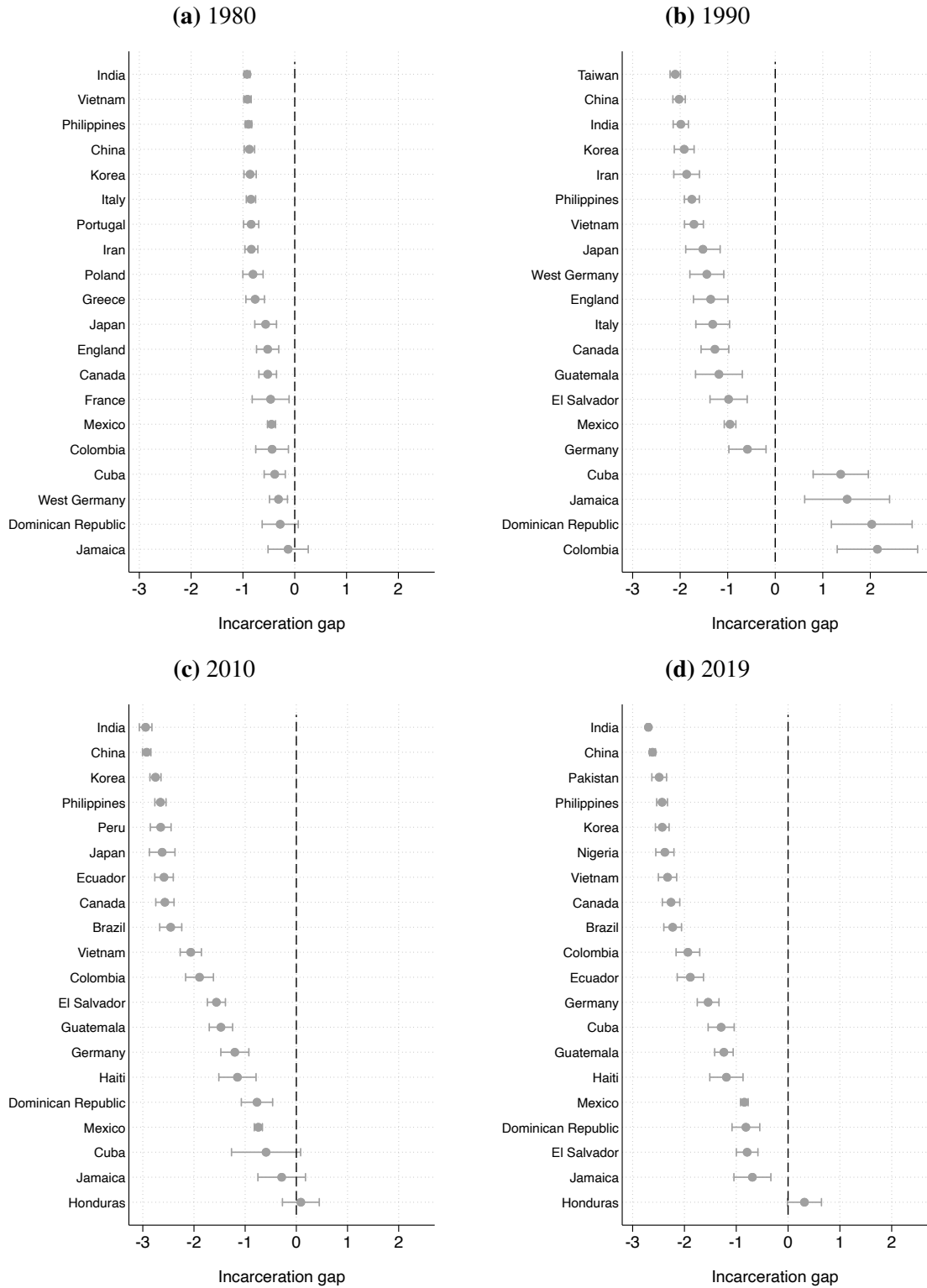
Notes: Panel (a) illustrates the share of men ages 18-40 that are foreign-born between 1870 and 2019. Panel (b) shows the composition of each immigrant group among foreign-born individuals. Each color depicts immigrants from a specific country-of-origin group, showing that immigrants today are more likely to come from Mexico and Central America as well as from the “rest of the world” group. For more details on the definition of each country-of-origin group, see Online Appendix B.

Figure A6: Incarceration Gap of Immigrants and US-born Men by Country of Origin, 1900-1930



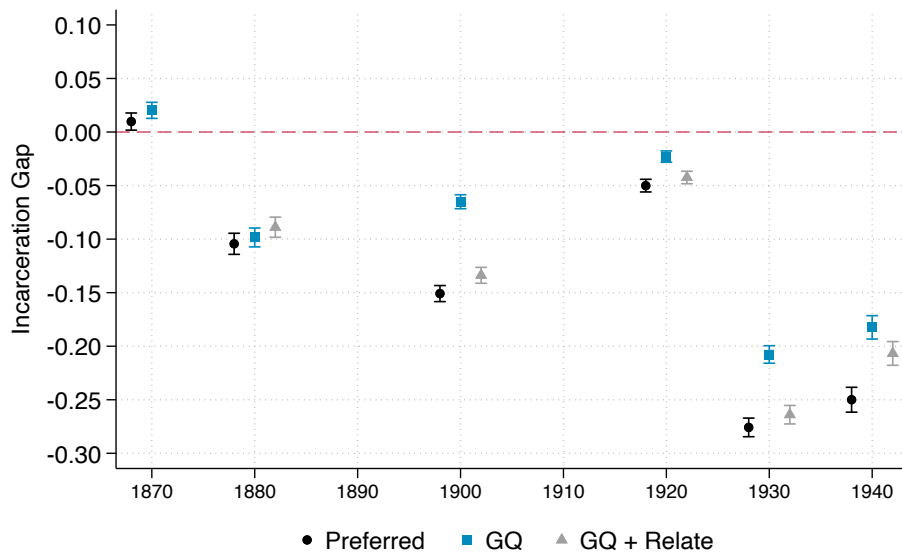
Notes: This figure plots the difference in incarceration propensities between US-born men and immigrants from each of the 20 sending countries with the largest populations in the US that year (each estimate is the value of β using equation (1) without any individual-level characteristics). All estimates report robust standard errors.

Figure A7: Incarceration Gap of Immigrants and US-born Men by Country of Origin, 1980-2019



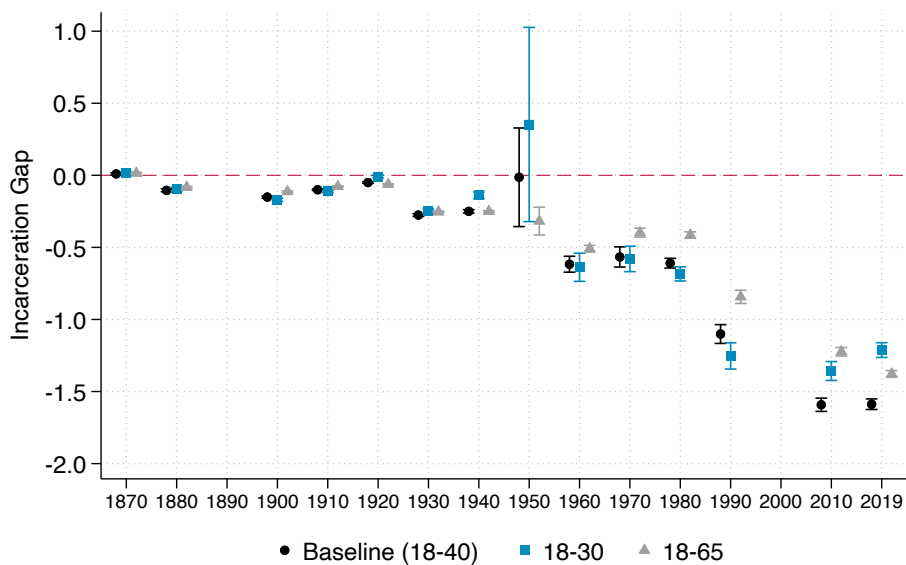
Notes: This figure plots the difference in incarceration propensities between US-born men and immigrants from each of the 20 sending countries with the largest populations in the US that year (each estimate is the value of β using equation (1) without any individual-level characteristics). All estimates report robust standard errors.

Figure A8: Incarceration Gap between Immigrants and US-born Men Using Alternative Incarceration Measures, 1870-1940



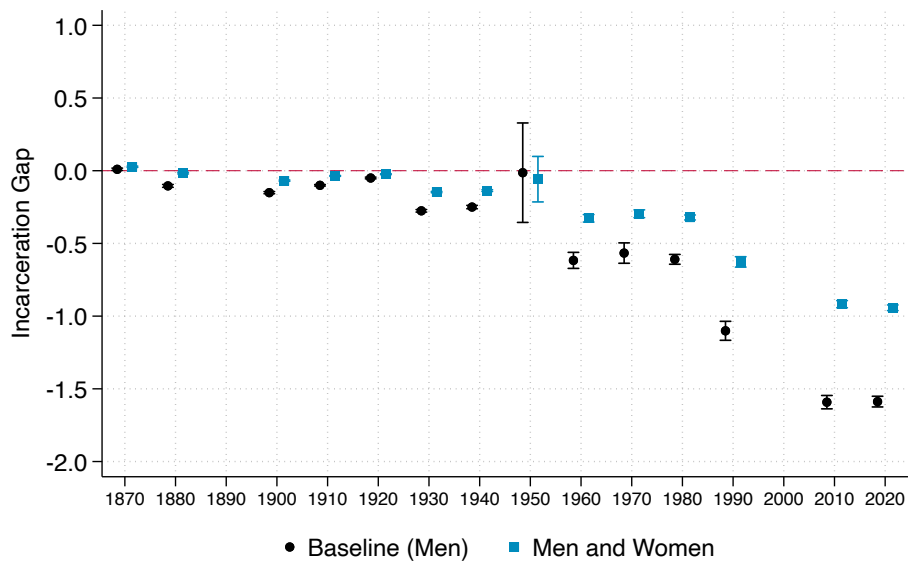
Notes: This figure plots the estimated values of β using equation (1) and varying the definition of incarceration in the 1870-1940 full-count decennial Censuses. The first series (labeled “Preferred”) utilizes the baseline measure of incarceration. The second series (labeled “GQ”) uses the IPUMS group quarters variable only to classify an individual as incarcerated. The third series (labeled “GQ + Relate”) uses the group quarters variable and the variable denoting an individual’s relationship to the household head to classify an individual as incarcerated. The 1910 Census does not identify group quarter types, so we omit this year in the comparison. The 1870 Census does not include a question on relationship to household head. For more details on these measures, see Online Appendix B. All estimates report robust standard errors.

Figure A9: Incarceration Gap between Immigrants and US-born Men, Varying the Age of the Sample



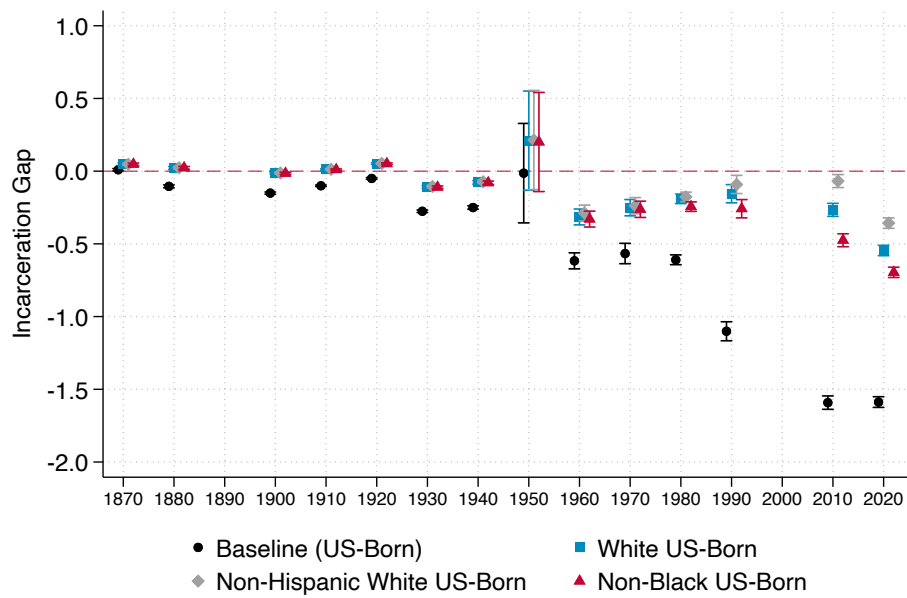
Notes: This figure plots the estimated values of β using equation (1) and varying the age of the individuals in the sample. The first series (labeled “Baseline”) reproduces the baseline estimates using men ages 18-40. The second and third series consider men ages 18-30 and 18-65, respectively. All estimates report robust standard errors.

Figure A10: Incarceration Gap between Immigrants and US-born Individuals, Including Women



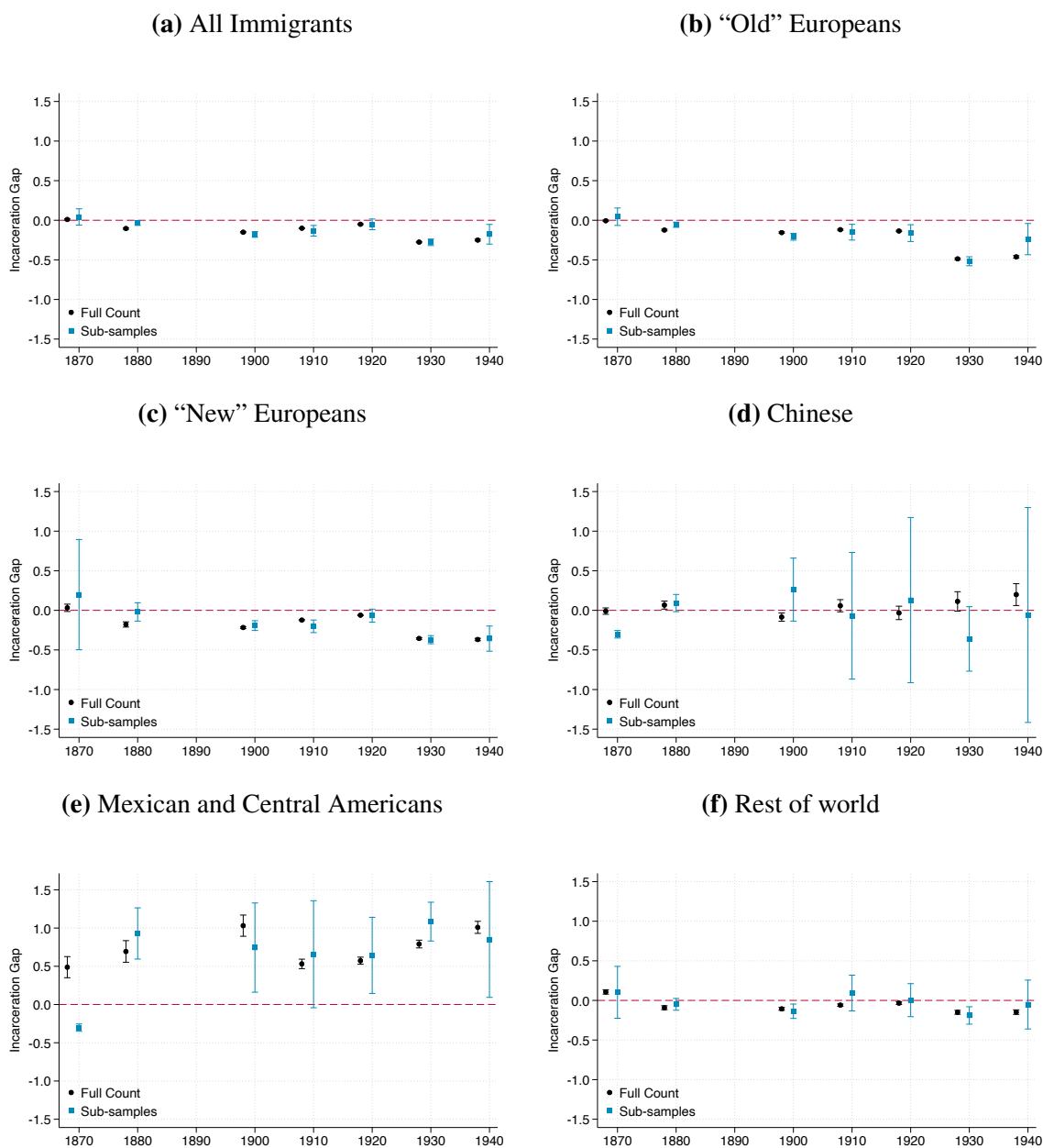
Notes: This figure plots the estimated values of β using equation (1) for individuals ages 18-40. The first series reproduces the baseline estimates restricting the sample to men. The second series expands the sample to include women. All estimates report robust standard errors.

Figure A11: Incarceration Gap between Immigrants and US-born Men, Using Alternative Groups of US-born Individuals



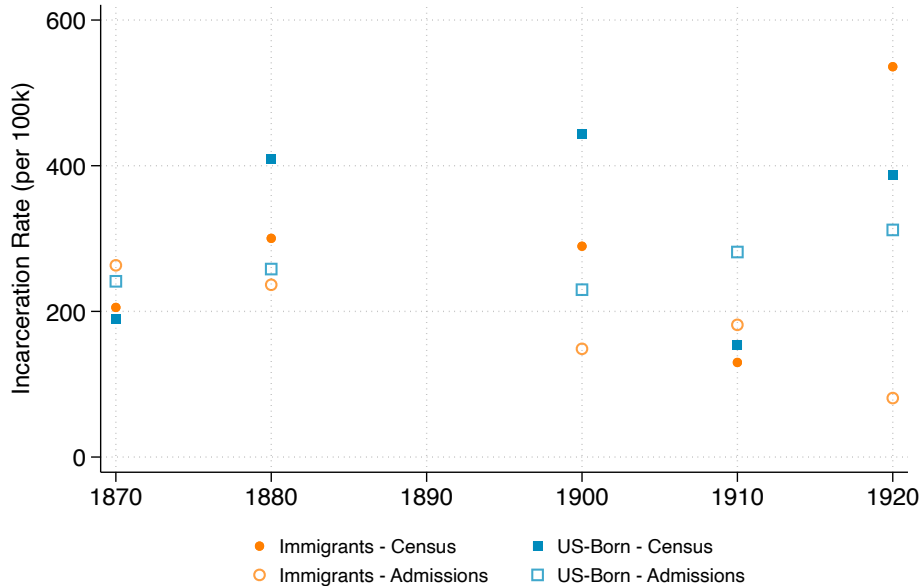
Notes: This figure plots the estimated values of β using equation (1) and varying the sample of US-born men. The first series reproduces the baseline estimate considering all US-born men. The second series only considers white US-born men. The third series considers non-Hispanic white US-born men. Hispanic individuals are identified using the “Hispan” variable provided by IPUMS. Before 1980, individuals were classified as Hispanic based on their country of birth, parental country of birth, Spanish surname, or relationship to someone identified as Hispanic through these characteristics. The fourth series considers US-born men whose race is not classified as Black. All estimates report robust standard errors.

Figure A12: Incarceration Gap between Immigrants and US-born Men, Comparing Full Count Census with Sub-samples, 1870-1940



Notes: This figure plots the estimated values of β using equation (1) for 1870-1940. The first series (labeled "Full Count") reproduces the baseline estimates using the full-count Censuses. The second series (labeled "Sub samples") utilizes the largest available sub-sample from each decennial Census. Panel (a) compares US-born men to all immigrants. Panels (b)-(f) compare US-born men to immigrants from a particular country-of-origin group. For more details, see the note to Figure 1 and Online Appendix B. All estimates report robust standard errors.

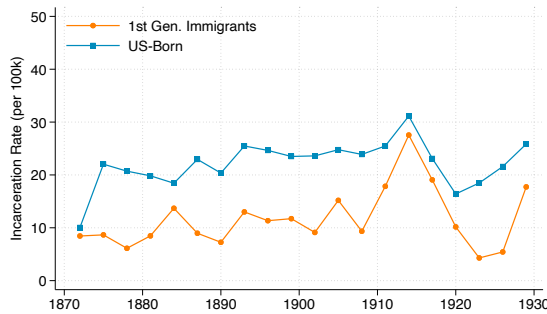
Figure A13: Comparison of Census-based Incarceration Rates in Missouri to Prison Admissions Rates from the Missouri State Penitentiary



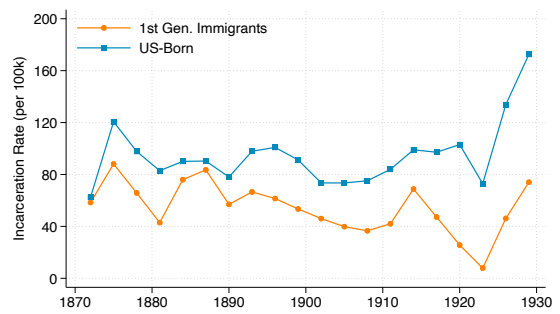
Notes: This figure compares the incarceration rates of immigrants and US-born men residing in Missouri (based on Census data) with prison admissions rates by nativity based on prison admission records from the Missouri State Penitentiary. The data on prison admissions come from digitized administrative records of the Missouri State Penitentiary, which covers the universe of prison inmates in Missouri. Population counts, used to calculate rates, come from the full-count Census.

Figure A14: Prison Admissions Rates of Immigrants and US-born Individuals in Missouri by Type of Crime, 1872–1929

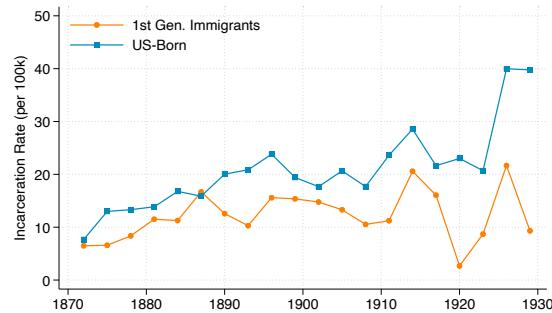
(a) Violent Crimes



(b) Property Crimes

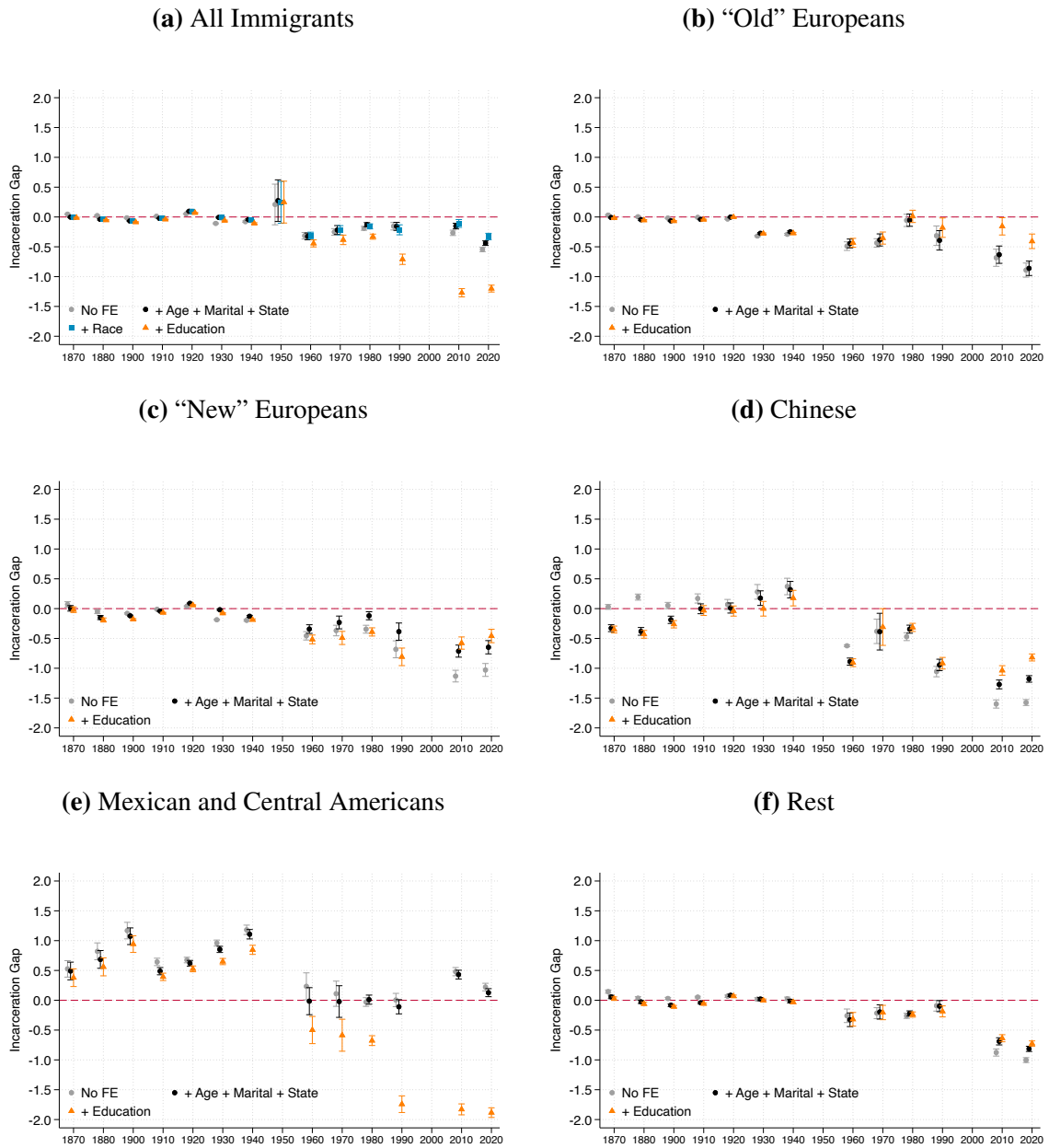


(c) Other Crimes



Notes: This figure plots prison admissions rates of immigrants and US-born individuals between 1872 and 1929 separately by crime type. Data are based on prison admission records from digitized administrative records of the Missouri State Penitentiary, which covers the universe of prison inmates in Missouri. Panels (a), (b), and (c) consider admissions for violent, property, and other crimes, respectively. Population counts, used to calculate rates, come from the full-count Census and are interpolated between Census years.

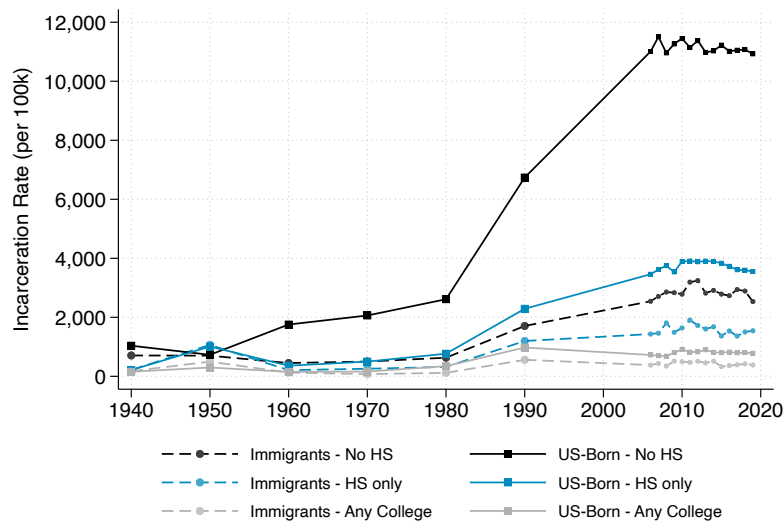
Figure A15: Difference in Incarceration Rates of Immigrants and White US-born Men, Adjusting for Individual-Level Characteristics, 1870-2019



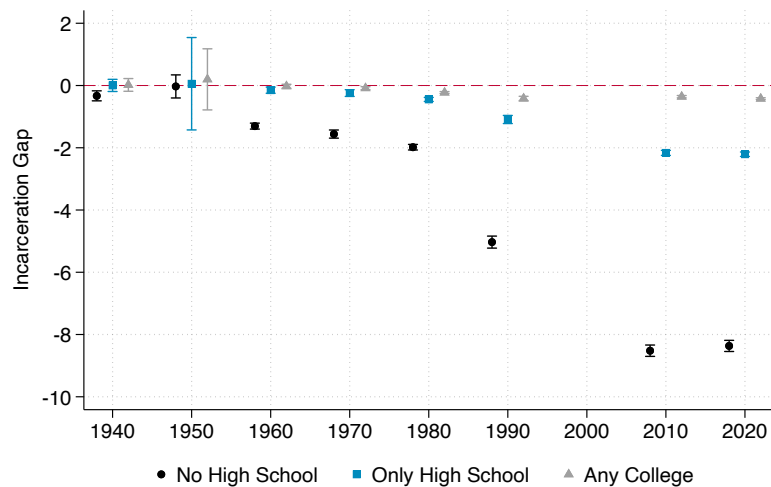
Notes: This figure is analogous to Figure 2 but restricts the sample of US-born men to white US-born men. For more details, see the note to Figure 2 and Online Appendix B. All estimates report robust standard errors.

Figure A16: Incarceration Gap Between Immigrants and US-born Men, by Educational Attainment, 1940-2019

(a) Incarceration Rates

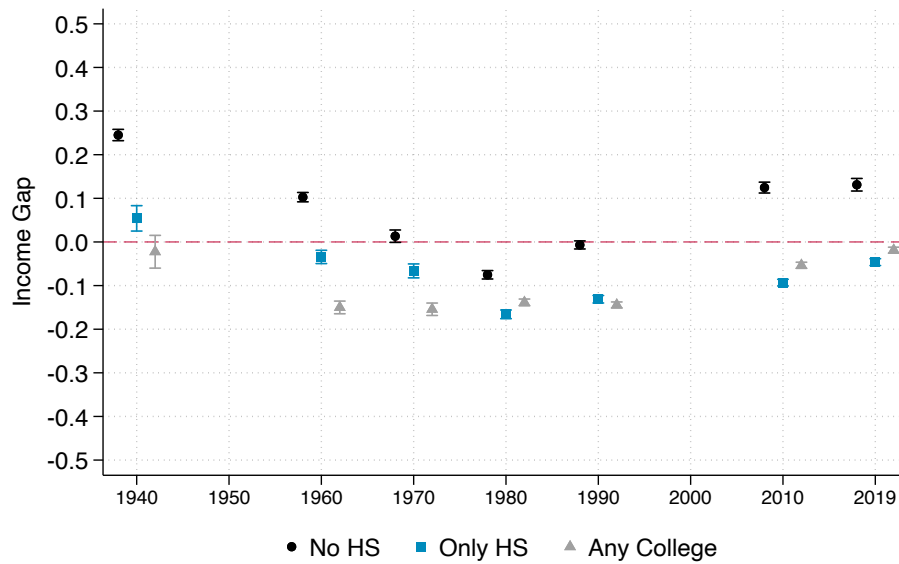


(b) Differences in Incarceration



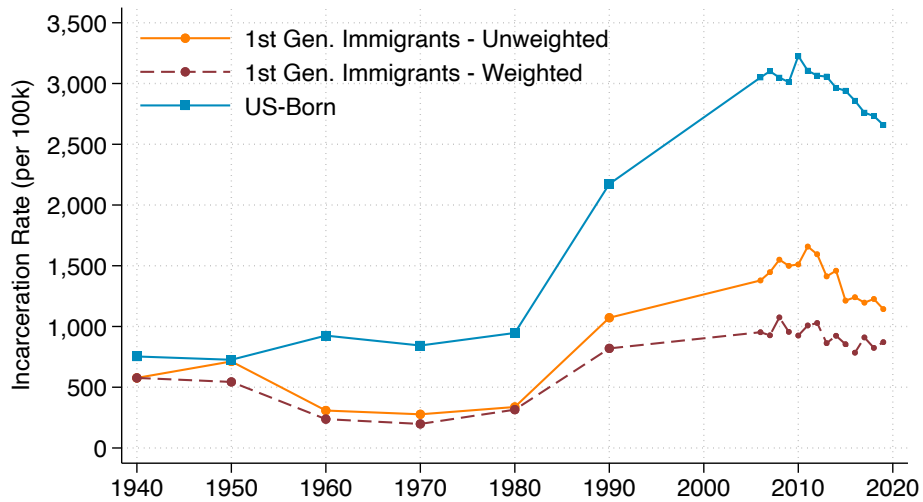
Notes: Panel (a) plots incarceration rates for immigrants and US-born men between 1940 and 2019 separately by educational attainment. Panel (b) plots the estimated values of β using equation (1) separately by individuals' educational attainment. "No High School" refers to individuals with 11 or fewer years of schooling. "High School" refers to individuals with exactly 12 years of schooling. "Any College" refers to individuals with one or more years of college. In panel (b), all estimates report robust standard errors.

Figure A17: Differences in Logged Income Between Immigrants and US-born Men, by Educational Attainment, 1940-2019



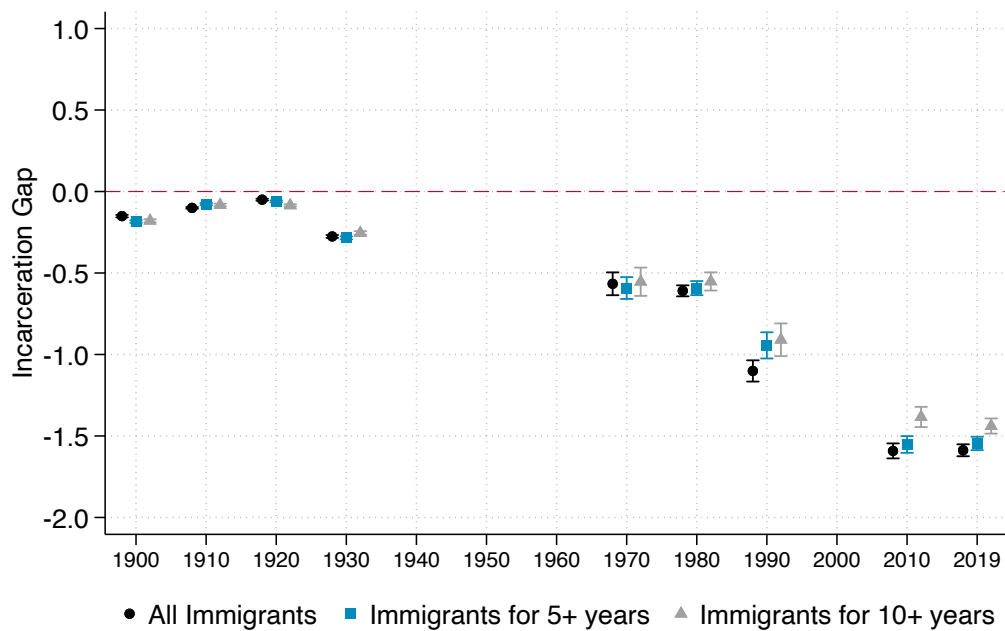
Notes: The figure plots the estimated values of β from equation (1) using logged income as the outcome variable and separately by individuals' educational attainment. The sample is restricted to men ages 18-40 who are in the labor force and have positive income. "No High School" refers to individuals with 11 or fewer years of schooling. "High School" refers to individuals with exactly 12 years of schooling. "Any College" refers to individuals with one or more years of college. All estimates report robust standard errors.

Figure A18: Incarceration Rate of Immigrants and US-born Men, Fixing the Immigrant Country-of-Origin Composition at 1940 Levels



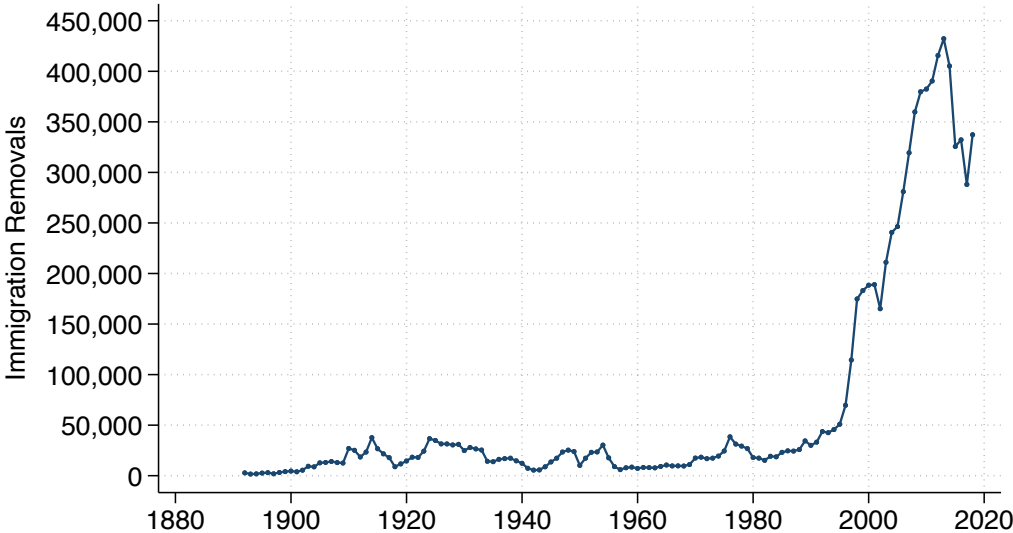
Notes: The first (orange) and third (blue) series plot the raw incarceration rates of immigrant men and US-born men, analogous to those in Figure 1. The second series (dashed red) holds fixed the immigrant composition in 1940 using the five country-of-origin groups (“old” Europeans, “new” Europeans, Chinese immigrants, Mexican and Central American immigrants, and immigrants from the “rest of the world”) and calculates the counterfactual incarceration rate after 1940 if each group’s incarceration had evolved naturally but their proportion in 1940 (as a share of all immigrants) remained fixed. This figure makes clear that if the immigrant composition had not changed since 1940, the immigrant incarceration rate would be lower than it actually is, and the immigrant-US-born incarceration gap would thus be even larger today.

Figure A19: Incarceration Gap between Immigrants and US-born Men, Excluding Recent Immigrants



Notes: This figure plots the estimated values of β from equation (1) and varying the sample of immigrants. The first series (labeled “All immigrants”) reproduces the baseline estimate including all immigrants regardless of time since arrival. The second and third series exclude individuals who arrived to the US within five and ten years, respectively. Estimates for 1940–1960 are omitted because the Census did not include a question about time since arrival to the United States in these years. All estimates report robust standard errors.

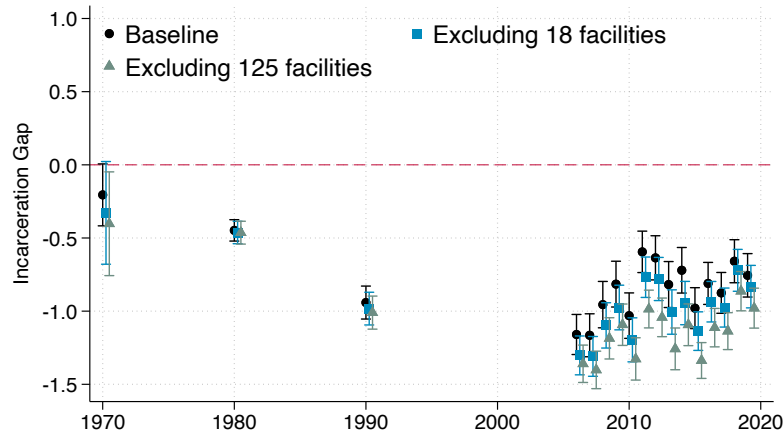
Figure A20: Number of Removals, 1892-2018



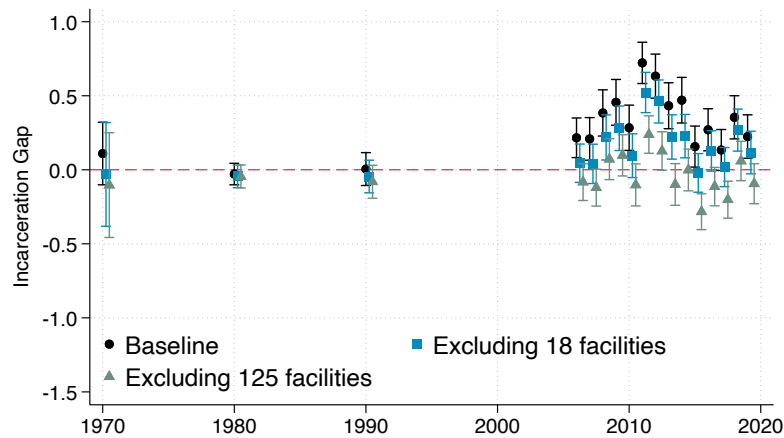
Notes: This figure plots the annual number of removals of inadmissible or deportable individuals between 1892 and 2018 using data from U.S. Department of Homeland Security (2018).

Figure A21: Incarceration Gap between Mexican and Central American Immigrants and US-born Men, Excluding Areas with ICE Facilities

(a) Relative to All US-born



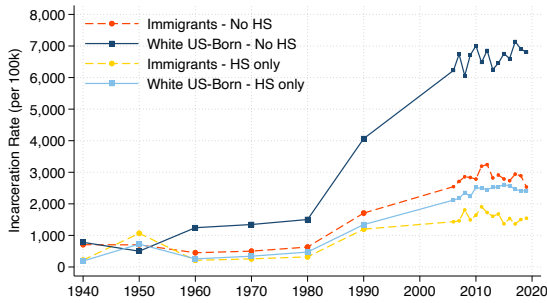
(b) Relative to White US-born



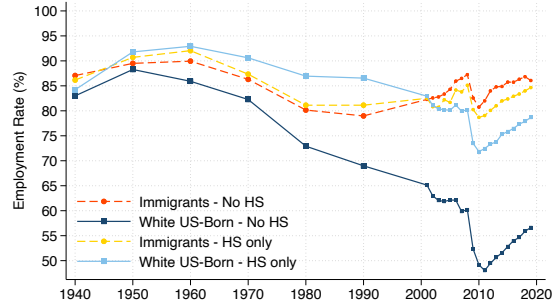
Notes: This figure plots the estimated values of β from equation (1) for Mexican and Central American immigrants and US-born men ages 18-40. Panel (a) compares these immigrants to all US-born men. Panel (b) restricts the comparison to white US-born men. The first series (labeled “Baseline”) in each panel uses the baseline sample. The second series (square markers) excludes the areas that included the 18 Immigration and Customs Enforcement (ICE) contract detention facilities and service processing centers as of 2022 (14-17 areas depending on the year). The third series (triangle markers) excludes the areas that included the 125 ICE contract detention facilities, service processing centers, facilities under intergovernmental service agreements, and US Marshall’s administered facilities as of 2022 (63-110 areas depending on the year). For more details on the areas excluded from the sample, see Online Appendix B. All estimates report robust standard errors.

Figure A22: Incarceration, Labor Market, Family Formation, and Health Outcomes of Immigrants and White US-born Men Without Any College Education, 1940-2019

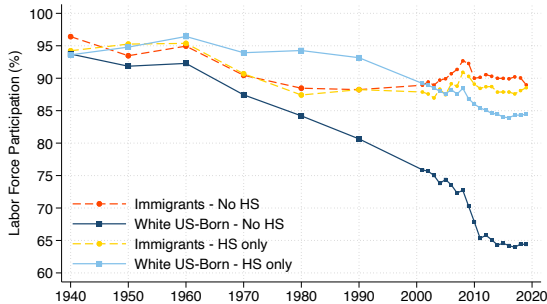
(a) Incarceration Rate (per 100,000)



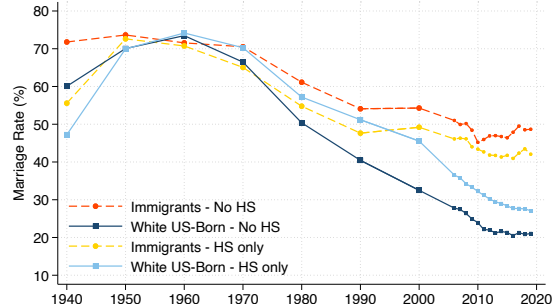
(b) Employment Rate (%)



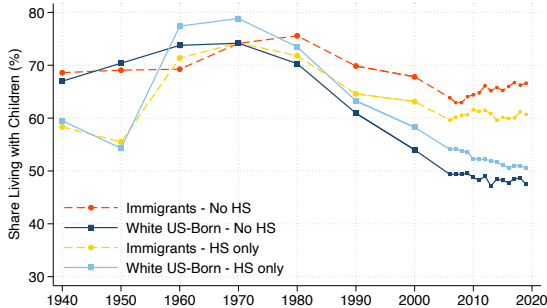
(c) Labor Force Participation Rate (%)



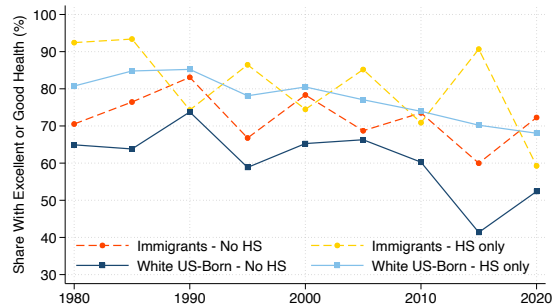
(d) Marriage Rate (%)



(e) Share Living with Children (%)



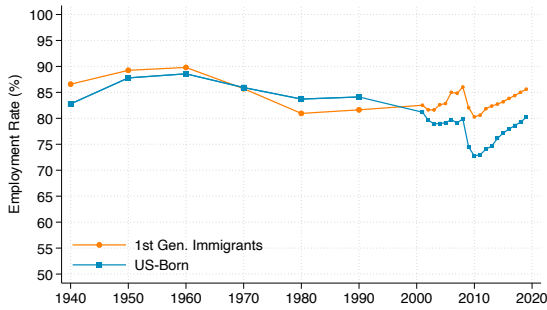
(f) Share with Excellent or Good Health (%)



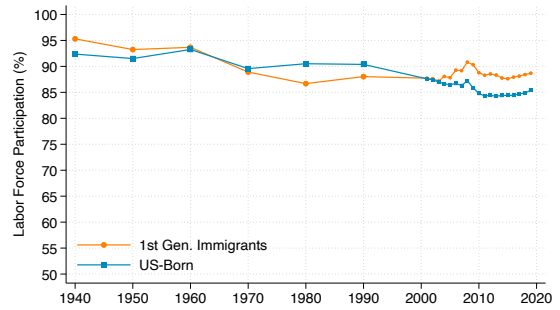
Notes: This figure is analogous to Figure 4 but restricts the sample of US-born men to white US-born men. For more details, see the note to Figure 4 and Online Appendix B.

Figure A23: Labor Market, Family Formation, and Health Outcomes of Immigrants and All US-born Men (Regardless of Educational Attainment), 1940-2019

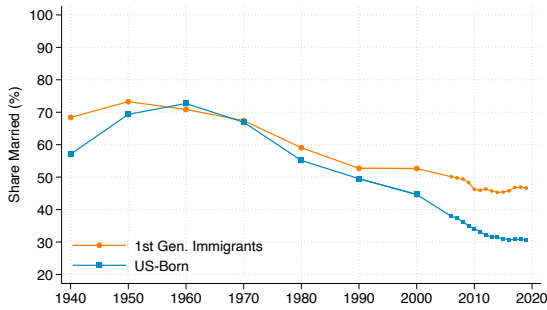
(a) Employment Rate (%)



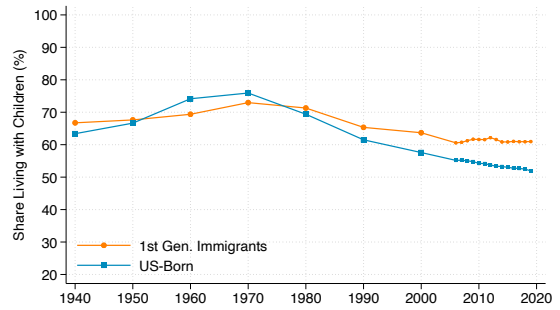
(b) Labor Force Participation Rate (%)



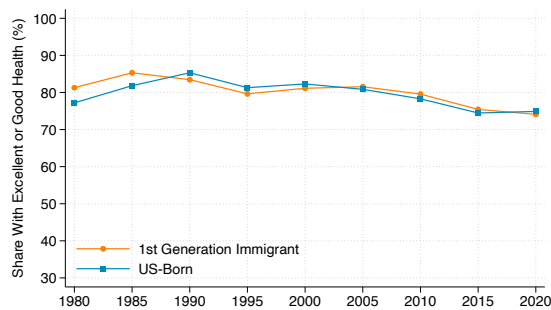
(c) Marriage Rate (%)



(d) Share Living with Children (%)



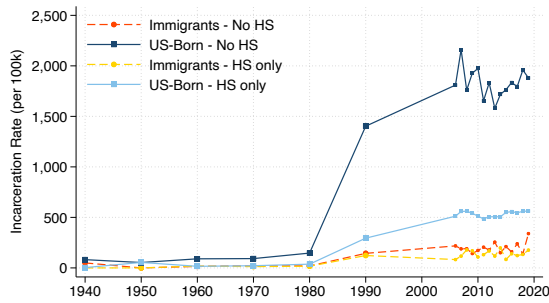
(e) Share with Excellent or Good Health (%)



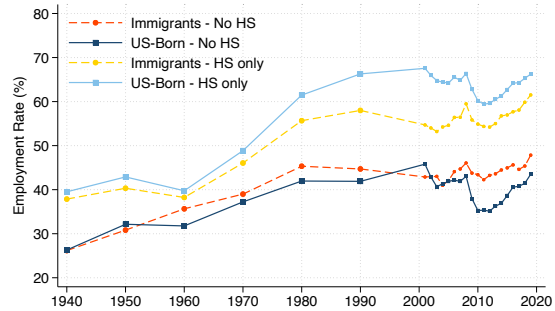
Notes: This figure is analogous to panels (b)-(f) of Figure 4 but considers all immigrants and US-born men regardless of educational attainment. For more details, see the note to Figure 4 and Online Appendix B.

Figure A24: Incarceration, Labor Market, Family Formation, and Health Outcomes of Immigrants and US-born Women Without Any College Education, 1940-2019

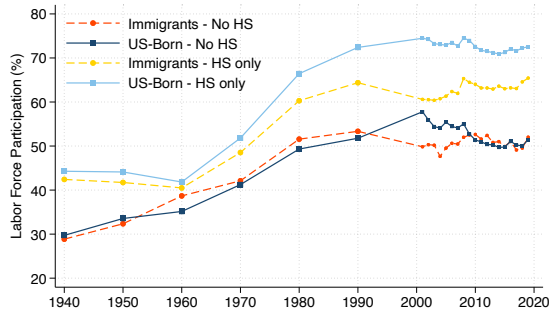
(a) Incarceration Rate (per 100,000)



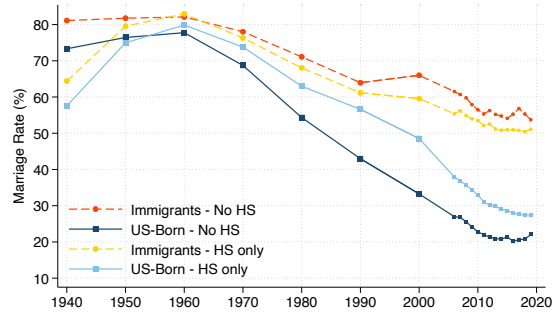
(b) Employment Rate (%)



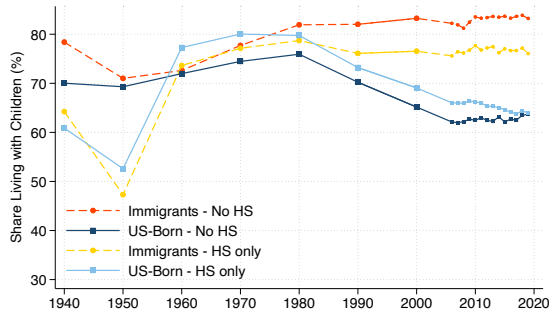
(c) Labor Force Participation Rate (%)



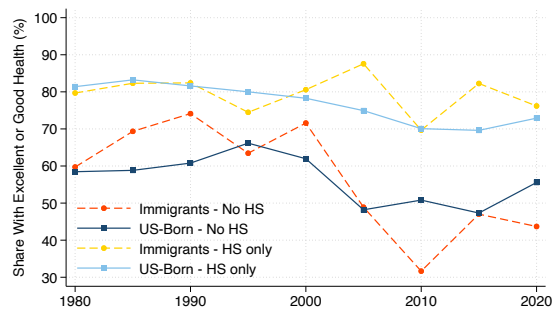
(d) Marriage Rate (%)



(e) Share Living with Children (%)

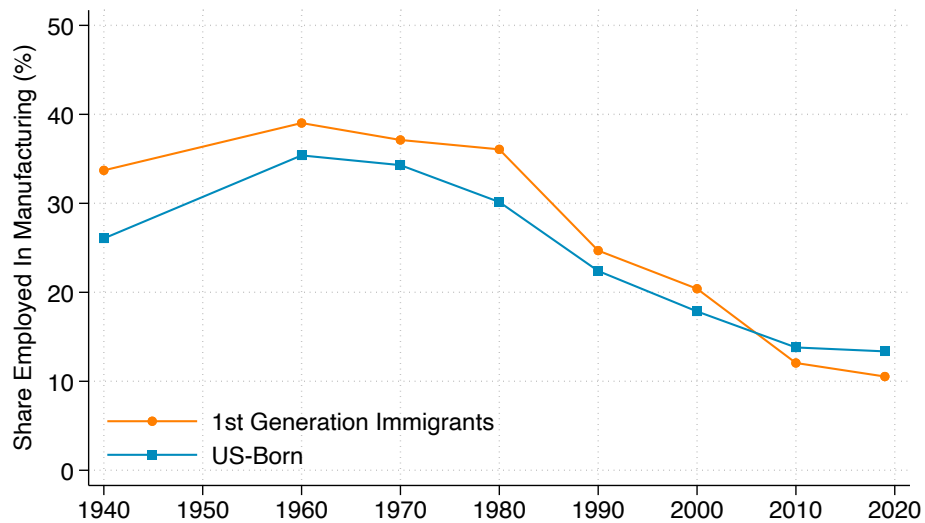


(f) Share with Excellent or Good Health (%)



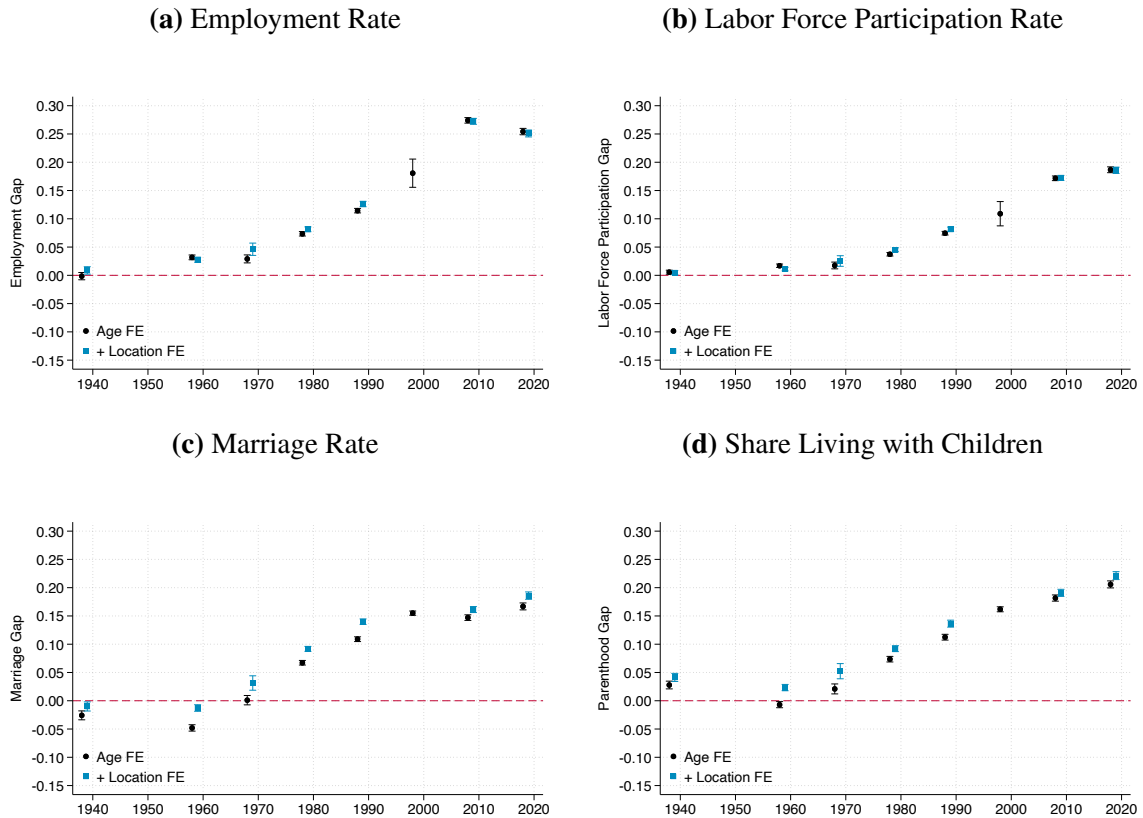
Notes: This figure is analogous to Figure 4 but expands the sample to include female immigrants and US-born individuals. For more details, see the note to Figure 4 and Online Appendix B.

Figure A25: Share of Low-Educated Immigrants and US-born Men Employed in Manufacturing



Notes: This figure plots the share of immigrants and US-born men ages 18-40 that were employed in manufacturing between 1940 and 2019. The sample is restricted to non-institutionalized men without a high school degree and who were in the labor force. This figure shows that the shares resembled each other until 2010, suggesting that compositional differences across declining industries cannot alone explain the immigrant-US-born differences in labor market outcomes (depicted in Figure 4).

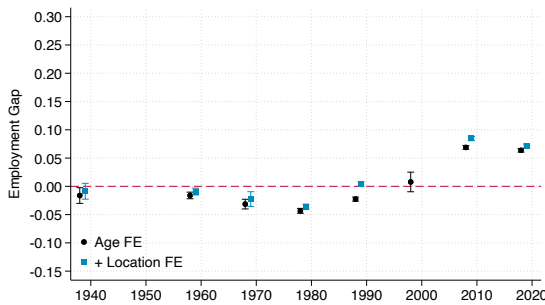
Figure A26: Differences in Labor Market and Family Formation Outcomes of Immigrants and US-born Men Without a High School Degree, Adjusting for Geography, 1940-2019



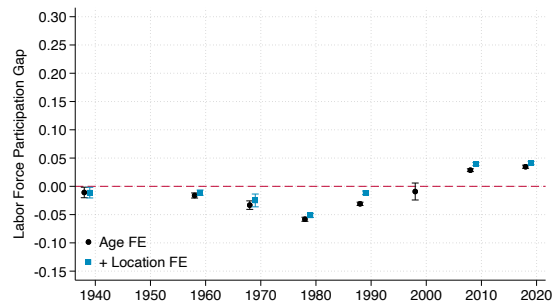
Notes: This figure plots the estimated values of β from equation (1) for immigrants and US-born men without a high school degree. Each panel considers a different outcome. The sample is non-institutionalized men ages 18-40 in panels (a)-(c) and ages 30-50 in panel (d). The first series (labeled “Age FE”) plots the estimated gaps including individual age fixed effects. The second series (labeled “+ Location FE”) adds location fixed effects. For 1940, we include county-of-residence fixed effects. For 1970 and 1980, we include fixed effects for each county group. For 1960 and 1990 onward, we include Public Use Metropolitan Area (PUMA) fixed effects. For more details, see the note to Figure 4 and Online Appendix B. All estimates report robust standard errors.

Figure A27: Differences in Labor Market and Family Formation Outcomes of Immigrants and US-born Men With Only a High School Degree, Adjusting for Geography, 1940-2019

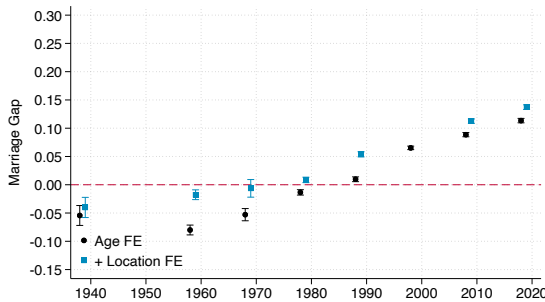
(a) Employment Rate



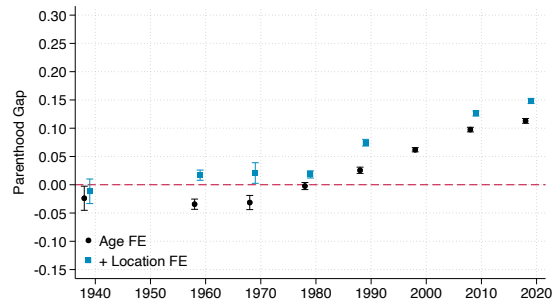
(b) Labor Force Participation Rate



(c) Marriage Rate



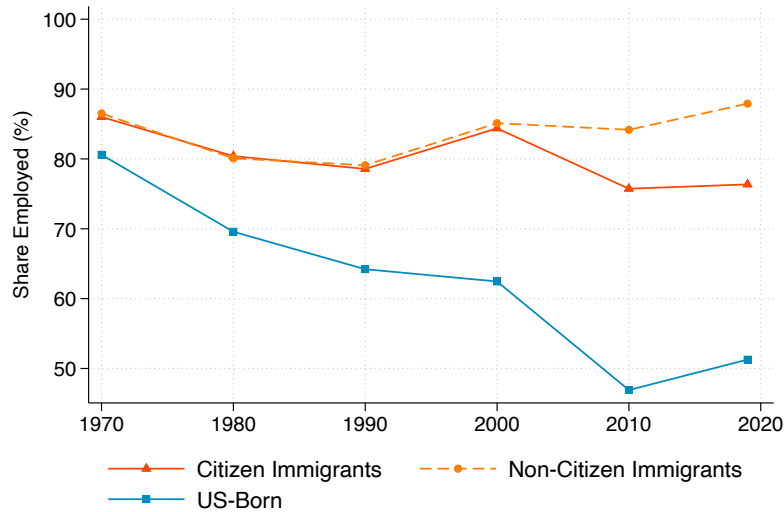
(d) Share Living with Children



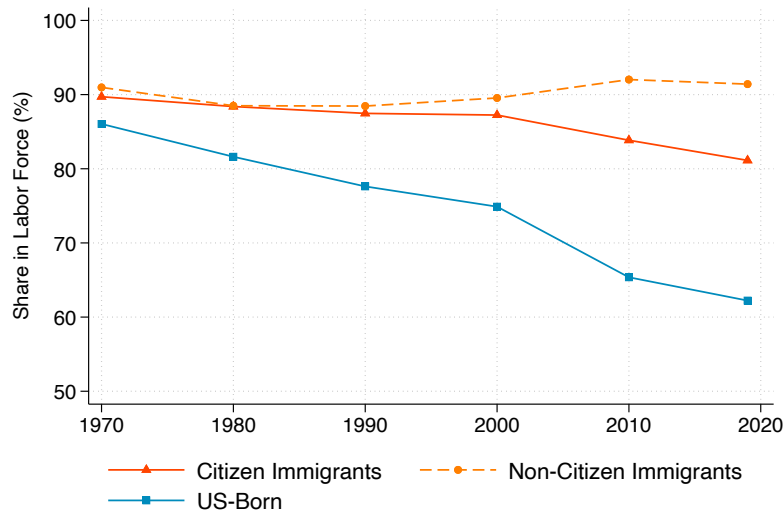
Notes: This figure plots the estimated values of β from equation (1) for immigrants and US-born men with only a high school degree. Each panel considers a different outcome. The sample is non-institutionalized men ages 18-40 in panels (a)-(c) and ages 30-50 in panel (d). The first series (labeled “Age FE”) plots the estimated gaps including individual age fixed effects. The second series (labeled “+ Location FE”) adds location fixed effects. For 1940, we include county-of-residence fixed effects. For 1970 and 1980, we include fixed effects for each county group. For 1960 and 1990 onward, we include Public Use Metropolitan Area (PUMA) fixed effects. For more details, see the note to Figure 4 and Online Appendix B. All estimates report robust standard errors.

Figure A28: Employment and Labor Force Participation Rates of Citizen and Non-Citizen Immigrants and US-born Men Without a High School Degree

(a) Employment Rate

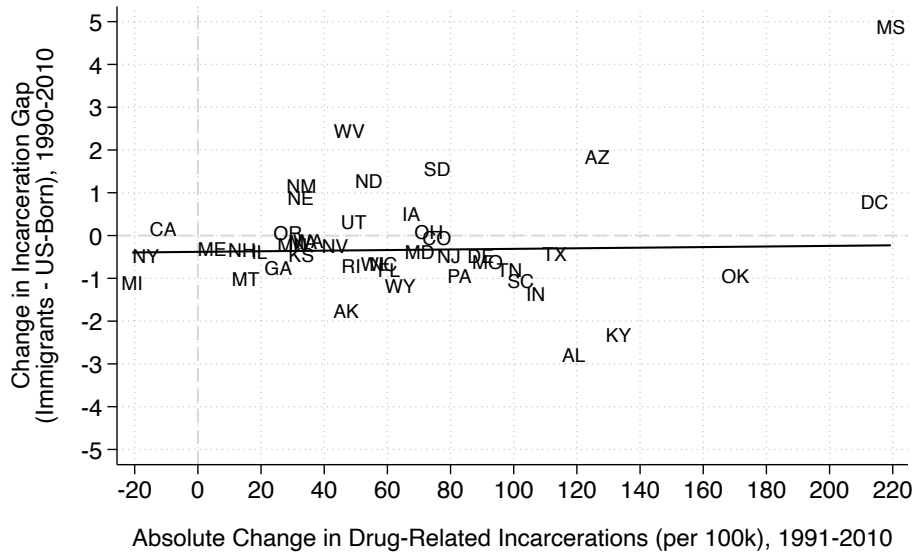


(b) Labor Force Participation Rate



Notes: This figure plots employment and labor force participation rates for citizen immigrants, non-citizen immigrants, and US-born men between 1970 and 2019. The sample is restricted to non-institutionalized men ages 18-40 who did not complete high school. Although the magnitude of the gaps between citizen migrants and the US-born are somewhat smaller in recent decades, the figure shows that less-educated citizen immigrants also have significantly higher employment and labor force participation rates than their US-born counterparts. It is thus unlikely that the availability of social insurance can explain the immigrant-US-born differences in labor market outcomes (depicted in Figure 4).

Figure A29: State-Level Changes in Drug-Related Incarcerations and the Immigrant-US-born Incarceration Gap Between 1990 and 2010



Notes: This figure plots state-level (absolute) changes in the incarceration rate of individuals convicted of drug-related offenses between 1991 and 2010 (x-axis) against changes in the immigrant-US-born incarceration gap, estimated using equation (1), in that same state and time period (y-axis). We use data from the National Corrections Reporting Program to calculate incarceration rates for drug-related offenses (averaging incarcerations between 1991 and 1993 and between 2008 and 2010 to calculate differences over this time period). This figure considers the potential role of drug crimes in explaining the widening of the immigrant-US-born incarceration gap: if US-born men are more likely to commit drug-related offenses and they are more likely to be incarcerated for these offenses than immigrants in the modern time period, then this difference could explain the relative decline in immigrants' incarceration rate. Put differently, if drug-related incarcerations are driving the increase, then we should find that the immigrant-US-born gaps are larger in states that experience large increases in drug-related incarcerations. This figure shows that at least when looking at state-level correlations, this does not seem to be the case.

Table A1: Overlap Between Alternative Incarceration Measures in the Full Count Censuses

	US-Born				Immigrants			
	GQ (1)	Preferred (2)	Both (3)	Share (%) (4)	GQ (5)	Preferred (6)	Both (7)	Share (%) (8)
1870	9,012	10,836	9,012	83	3,174	3,573	3,174	89
1880	28,613	34,615	28,262	82	5,006	6,322	4,970	79
1900	35,904	53,626	33,748	63	6,788	8,623	6,554	76
1910	–	43,631	–	–	–	8,165	–	–
1920	38,689	51,132	36,949	72	7,829	9,624	7,561	79
1930	125,993	149,380	122,197	82	13,077	14,609	12,672	87
1940	126,576	165,699	57,691	35	4,758	6,826	2,320	34

Notes: This table shows the number of incarcerated individuals in each Census year separately by nativity and by measure of incarceration. “GQ” refers to the number of men classified as incarcerated using the IPUMS group quarters variable. “Preferred” refers to the number of men classified as incarcerated using our preferred measure that combines information from the group quarters variable with the original strings of the “group quarters,” “occupation,” and “relationship to household head” variables. “Both” refers to the number of men classified as incarcerated under both approaches. “Share” refers to the share of incarcerated men under the preferred measure that would have also been classified as incarcerated using only using the group quarters variable (column 3 divided by column 2 and column 7 divided by column 6). For more details, see Online Appendix B.

Table A2: Sample Size for Immigrants, by Year and Country Group

Inc. (1)	US-born			All Immigrants			"Old" Europeans			"New" Europeans			Chinese			Mexican & Central Americans			Rest of the world		
	Total (2)	Inc. (3)	Inc. Rate (4)	Total (5)	Inc. (6)	Inc. Rate (7)	Total (8)	Inc. (9)	Inc. Rate (10)	Total (11)	Inc. (12)	Inc. Rate (13)	Total (14)	Inc. (15)	Inc. Rate (16)	Total (17)	Inc. (18)	Inc. Rate (19)	Total (20)	Inc. (21)	Inc. Rate (21)
1870	10,836	5,299,875	204	3,573	1,667,878	214	2,792	1,406,536	199	42,743	236	91	46,680	195	96	13,855	693	493	158,064	312	
1880	34,615	7,625,747	454	6,322	1,808,660	350	4,572	1,383,229	331	98,575	276	404	77,618	520	251	21,868	1,148	823	227,370	362	
1900	53,626	11,761,318	456	8,623	2,826,309	305	4,810	1,600,221	301	737,100	239	196	52,810	371	435	29,255	1,487	1,421	406,923	349	
1910	43,631	14,574,042	299	8,165	4,101,636	199	2,621	1,451,220	181	2,095,851	176	83	23,196	358	687	82,717	831	1,077	448,652	240	
1920	51,132	16,339,910	313	9,624	3,661,154	263	1,659	934,362	178	2,228,389	252	42	14,975	280	1,395	157,491	886	910	325,937	279	
1930	149,380	19,709,041	758	14,609	3,030,274	482	2,261	836,267	270	1,558,613	403	192	22,048	871	3,537	228,384	1,549	2,341	384,962	608	
1940	165,699	23,081,996	718	6,826	1,458,866	468	1,166	456,094	256	6,075,595	350	168	18,327	917	1,819	105,338	1,727	1,549	271,512	571	
1950	556	302,177	726	17	8,946	712	3	2,085	523	3,162	246	1	226	1,334	10	1,239	3,346	1	2,234	160	
1960	11,515	1,244,704	925	132	42,800	308	14	10,393	135	13,746	167	0	1,237	0	55	6,418	857	40	11,006	363	
1970	7,179	851,088	844	103	37,146	277	7	7,511	93	8,083	161	2	1,358	147	35	5,484	638	46	14,710	313	
1980	17,992	1,900,112	947	461	136,617	337	81	17,182	471	16,429	183	3	5,243	57	185	37,082	499	162	60,681	267	
1990	29,169	1,984,280	2,173	1,909	229,569	1,072	157	19,745	913	14,786	545	18	9,960	172	750	89,747	1,232	912	95,331	1,136	
2000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2010	90,995	1,758,597	3,165	10,336	340,376	1,574	329	17,368	1,156	20,280	708	70	16,146	241	7,185	149,787	2,319	2,489	136,795	964	
2019	97,028	1,892,429	2,790	8,284	326,127	1,203	263	17,020	857	21,283	720	79	22,060	174	5,402	116,614	1,970	2,276	149,150	745	

Notes: This table presents the sample size and incarceration rates for US-born men and for immigrant men from different country groups. The sample is restricted to men ages 18–40. Whenever available, we use sample weights provided by IPUMS to calculate incarceration rates (per 100,000). For more information about each data source, see Online Appendix B.

Table A3: Characteristics of Immigrants and US-Born Men, by Incarceration Status and Period

	1870–1940						1940–1970						1980–2019					
	US-Born			Immigrants			US-Born			Immigrants			US-Born			Immigrants		
	Incarcerated (1)	Not Incarcerated (2)	(3)	Incarcerated (4)	Not Incarcerated (5)	(6)	Incarcerated (7)	Not Incarcerated (8)	(9)	Incarcerated (10)	Not Incarcerated (11)	(12)	Incarcerated (13)	Not Incarcerated (14)	(15)	Incarcerated (16)	Not Incarcerated (17)	(18)
Age	26.9	27.9	29.3	30.5	27.2	28.3	28.5	30.5	29.2	28.4	29.9	30.3	29.2	28.4	29.9	28.4	29.9	30.3
Married	28.8	51.6	26.8	56.3	31.0	63.2	33.4	65.3	14.3	34.6	24.4	46.9	14.3	34.6	24.4	34.6	24.4	46.9
Region																		
North	20.8	24.5	44.6	47.9	18.4	23.5	26.6	41.3	13.9	16.9	13.8	19.8	13.9	16.9	13.8	16.9	13.8	19.8
Midwest	29.0	33.2	26.6	33.4	23.3	28.3	13.1	20.1	20.1	23.3	9.0	12.4	20.1	23.3	9.0	23.3	9.0	12.4
South	41.7	34.7	9.8	5.8	39.6	32.2	28.7	13.3	45.6	37.1	43.0	33.5	45.6	37.1	43.0	37.1	43.0	33.5
West	8.4	7.5	18.9	12.9	18.7	16.0	31.6	25.3	20.4	22.7	34.1	34.3	20.4	22.7	34.1	22.7	34.1	34.3
Education																		
Illiterate	15.6	7.5	16.1	11.9
Literate	84.4	92.5	83.9	88.1
No High School	77.8	45.2	78.7	52.3	36.3	9.6	50.6	25.4	36.3	9.6	50.6	9.6	50.6	25.4
Only High School	17.3	30.2	13.9	19.6	49.8	40.1	35.9	31.6	49.8	40.1	35.9	40.1	35.9	31.6
Any College	4.9	24.6	7.4	28.0	13.9	50.3	13.5	43.0	13.9	50.3	13.5	50.3	13.5	43.0
Race																		
White	63.5	87.3	94.4	96.5	58.5	89.1	84.0	88.7	45.0	77.8	50.2	46.2	45.0	77.8	50.2	77.8	50.2	46.2
Black	36.0	12.5	2.5	0.6	39.8	10.1	8.2	3.2	43.8	13.2	8.1	8.1	43.8	13.2	8.1	13.2	8.1	8.1
Other	0.5	0.3	3.1	2.8	1.6	0.8	7.7	8.1	11.1	9.0	38.1	45.7	11.1	9.0	38.1	9.0	38.1	45.7
Observations	343,220	74,966,713	50,916	17,044,995	1,301,780	156,554,705	27,550	6,891,281	19,443,177	666,781,639	1,880,911	140,088,216	19,443,177	666,781,639	1,880,911	666,781,639	1,880,911	140,088,216

Notes: This table presents characteristics of immigrant and US-born men by incarceration status for three sample periods. The sample is restricted to men ages 18–40. Whenever available, we use sample weights provided by IPUMS to calculate averages for each variable. “Observations” refers to the unweighted number of observations. For more information about each data source, see Online Appendix B.

ONLINE APPENDIX B: ADDITIONAL DETAIL ON DATA SOURCES

Data Sources: Census and ACS

We combine the full-count decennial Censuses between 1870 and 1940 (excluding 1890) with the largest available subsample of each Census between 1950 and 2000 and the American Community Survey for the more recent period. We recover the full-count decennial Censuses from the IPUMS datasets in the NBER server (Ruggles et al. 2021) and the Census subsamples and the ACS from the IPUMS website (Ruggles et al. 2022). In particular, we use the following:

- 1870, 1880, 1900, 1910, 1920, 1930, and 1940 full-count decennial Censuses.¹
- 1950 1% weighted sample
- 1960 5% unweighted (flat) sample
- 1970 pooled 1% FORM 1 unweighted state, metro and neighborhood samples. Form 1 compiles a set of variables that were asked to 5% of the population, which is included in these samples
- 1980 5% unweighted (flat) state sample
- 1990 5% weighted state sample
- 2000 5% weighted state sample
- 2001–2005 annual ACS weighted sample corresponding to up to 1% of the population (depending on the year). Excludes individuals in group quarters.
- 2005–2019 annual ACS weighted sample corresponding to 1% of the population in each year
- 2008–2012 5-year ACS weighted sample corresponding to 5% of the population
- 2015–2019 5-year ACS weighted sample corresponding to 5% of the population

We also collect historical subsamples from IPUMS for robustness exercises:

- 1870 1% unweighted (flat) sample
- 1880 10% weighted sample
- 1900 5% unweighted (flat) sample
- 1910 1% unweighted (flat) sample
- 1920 1% unweighted (flat) sample
- 1930 5% unweighted (flat) sample
- 1940 1% weighted sample

¹ For 1870–1940, we use the full-count Census files located in the following directory of the NBER server: /home/data/census-ipums/v2021/dta/. For 1940, we use the file located in /homes/data/cens1940/20180316/100files/ to produce alternative measures of incarceration (i.e., our “GQ” and “Relate” measures, as described in this appendix).

We use annual ACS samples to plot incarceration rates and five-year samples to estimate differences in incarceration between immigrants and the US-born. We do not pool annual and five-year samples for the same analysis.

Our baseline results restrict the sample to men ages 18–40. Given its small sample size, we exclude the 1950 Census from results that split immigrants by country-of-origin group. Throughout the analysis, we utilize person weights provided by IPUMS.

Defining US-born, immigrants, and country groups

We define immigrants as individuals who were not born in any US state or outlying US area or territory. The US-born includes every individual not coded as an immigrant under this definition. Following Butcher and Piehl (2007), we exclude from the sample individuals born in outlying areas of the United States as well as those born abroad to US citizens.

We define the following five countries-of-origin groups for immigrants:

- “Old Europeans”: individuals born in the countries that belong to Northern and Western Europe including Germany (IPUMS codes 400–429 and 453).
- “New Europeans”: individuals born in the countries that belong to Southern Europe, Central/Eastern Europe, and the former USSR (IPUMS codes 430–499 excluding 453).
- Individuals born in China.
- Individuals born in Mexico and Central America.
- “Rest of the World”: individuals born in other countries in Asia, Africa, Oceania, the Caribbean, and South America.

Measuring incarceration

Full-count censuses

Incarceration can in principle be measured in the full-count data using the “group quarters” and “group quarter type” variables available from the Census. Prisoners are defined as those who reside in institutional and other group quarters and whose group quarter type corresponds to correctional institutions. Correctional institutions include federal and state correctional facilities, prisons, penitentiaries, military prisons, local correctional facilities, jails, school juvenile delinquents, reformatory, camp or chain gangs, and houses of correction.

However, these variables were not consistently coded to identify prisoners in the full-count Census data (see Eriksson 2020 for a discussion).² Common issues with these variables involve individuals who were not incarcerated but were counted as such, individuals that were actually incarcerated but appeared in households, and individuals that lived in prisons but were not incarcerated (such as prison guards). An additional issue is the classification of individuals defined solely as inmates, who may not be incarcerated in a correctional facility (e.g., inmates who frequent or live in mental and elderly institutions or those in non-institutional group quarters).

² Eriksson (2020) implements a different classification procedure from us. Specifically, that paper uses the 1920–1940 full-Count Census along with images looked up by hand to classify individuals born in fourteen states in the US South as incarcerated.

To account for these issues, we construct our preferred incarceration measure for the full-count Census data using the following procedure:

1. For each individual in the data, we retrieve their “group quarters,” “group quarter type,” “relate,” and “occupation” variables (i.e., the code as well as the original strings as reported in the Census).
2. Next, we define individuals as incarcerated using information in the “relate” string variable if they meet *any* the following requirements:
 - a. *Explicit correctional string*: Individuals who have the following words and their spelling variations in the “relate” string variable: “Prisoner,” “Convict,” or “Jail.” At this step, we exclude individuals whose “relate” string variable conveys a relationship to “Prisoner,” “Convict,” or “Jail,” such as “Daughter,” “Son,” “Wife,” “Head,” as well as “Guard,” “Jailer,” “Chief,” “Helper,” “Officer,” “Manager,” “Charge,” “Superintendent,” including their spelling variations. (i.e., we exclude an individual whose “relate” string variable is “Prisoner guard,” “Convict daughter,” etc.).
 - b. *Inmate and explicit correctional institution string*: Individuals who have the following words and their spelling variations combined with the word “Inmate” in the “relate” string variable: “Prison,” “Jail,” “Penitentiary,” “Reformatory,” and “Correction.” We exclude individuals classified by the “group quarter type” variable as part of a mental institution, an institution for the elderly, handicapped, and poor, or a non-institutional group quarter. This avoids counting individuals who reside in these institutions as inmates, but for whom it is not clear that they are serving a criminal sentence.
 - c. *Inmate with missing information in the string variable*: Individuals who have the word “Inmate” (without any additional words) in the “relate” string variable or who have a missing value, an “X,” or a “*” in the “relate” string variable. These individuals are classified as incarcerated if either:
 - i. their “group quarters” string variable contains the words “Prison,” “Jail,” “Penitentiary,” “Reformatory,” “Correction,” “Convict,” “Delinquent,” “Penal,” and other grammatical variations of these words; or
 - ii. their “group quarters type” variable code corresponds to a correctional institution when the relate string says “Inmate.” For individuals with missing values, “X,” or “*” in the relate string variable, we additionally condition on whether the individual is an institutional inmate based on their “relate” variable code.
3. We follow the steps in (2) to classify individuals as incarcerated using the “occupation” string variable.
 - a. We follow the procedure in (2.a) (i.e., an individual is identified as incarcerated if their occupation includes “Prisoner,” “Convict,” or “Jail.”). Because the “occupation” string does not convey familial relationships, we do not exclude any individuals in this step based on their relationship to household. However, we do exclude individuals if their occupation denotes a potential non-prisoner occupation (“Guard,” “Jailer,” “Chief,” “Helper,” “Officer,” “Manager,” “Charge,” and “Superintendent”).
 - b. We replicate step (2.b) exactly.

- c. We replicate step (2.c), but in addition to “Inmate,” “X,” and “*,” we also include individuals in this step whose occupation string variable says: “No Occupation,” “No,” “None,” “Without Occupation,” “Nothing,” or has a missing value.³

In our preferred measure of incarceration, we define an individual as incarcerated if they are classified as such in steps one through three.⁴

The 1870 Census does not include the “relate” string variable. We classify individuals as incarcerated in these years using the “occupational” string variable (step 3). In addition, we include individuals as incarcerated if their “relate” variable code is “institutional inmate” and their “group quarter type” variable code corresponds to correctional institutions.

The 1910 Census does not identify group quarter types. In this case, we rely on our preferred measure to classify prisoners based on strings of the “relate” and “occupation” variables that clearly identify individuals as prisoners (as in step 2.a). However, due to the lack of the “group quarter” string variable and the “group quarter type” variable, we are unable to implement steps 2.b, 2.c., 3.b, and 3.c.

For robustness checks, we also construct two alternative measures of incarceration, which we refer to as the “GQ measure” and the “relate measure.” The “GQ measure” refers to individuals who reside in institutional and other group quarters and whose group quarter type corresponds to correctional institutions (without any additional modifications). The “relate measure” refers to individuals who satisfy the “GQ measure” and either steps (2.a) or (2.b). In the “relate measure,” we exclude individuals who appear to be incarcerated via the “GQ measure,” but who are coded as family members of the household head in their “relate” variable code.

We note that the paper’s main takeaways are similar when using just IPUMS group quarters variable, rather than this more detailed approach.

Census subsamples and ACS

Between 1950 and 1980, we define prisoners as those who belong to institutional and other group quarters and whose group quarter type corresponds to correctional institutions (analogous to the GQ measure described above). For 1910, group quarter types were imputed by IPUMS. Between 1990 and 2019, the “group quarter” variables only allow us to identify institutionalized individuals, but not those who are institutionalized in adult correctional facilities. In this case, we identify incarcerated individuals as those who are classified as living in institutional group quarters.

³ To be conservative, when an individual is classified as incarcerated using missing information under the relate string (step 2c), but not under the occupation string (step 3c), we only identify an individual as incarcerated if they are classified as institutional inmates in their “relate” variable code or if their “relate” variable string is the word “Inmate.”

⁴ The 1940 Census presents a comparability issue among large households. According to IPUMS: “Before 1940 and in 1980–1990, units with 10 or more individuals unrelated to the householder are considered group quarters.” We adjust our “preferred” measure in 1940 to include individuals whose “relate” variable string says “Inmate” (in cases where the “group quarters” variable code is “Other Group Quarters” and the “group quarter type” variable code indicates a “Non-group quarter household”). For more details, see https://usa.ipums.org/usa-action/variables/GQ#comparability_section.

Other variable definitions

Education

We use the “education” variable in each sample to assign individuals to three educational groups: high school dropouts (i.e., those with no schooling up to those who completed grade 11), high school only (grade 12), and any college (1 or more years of college). These three groups comprise the educational fixed effects used in our analysis. This variable is defined starting with the 1940 Census.

Race

We use the “race” variable in each sample to assign individuals to three racial groups: white, Black, and “other” (referring to individuals whose race classification is neither white nor Black). These three groups comprise the race fixed effects used in our analysis.

Marital status

We use the “marital status” variable in each sample to assign individuals to three marital status groups: married (married, spouse present or absent); separated, divorced, or widowed; and never married/single. These three groups comprise the marital status fixed effects used in our analysis, and we use the married category to construct marriage rates. This variable is defined for every year.

State of residence

To compare individuals living in similar geographies, we use state-of-residence fixed effects. Although most individuals convicted of crimes are incarcerated in their state of residence, we cannot control for geography below the state level because inmates can be incarcerated in correctional facilities far from their initial residential location (i.e., their county of residence at the time of the Census may not reflect their county of residence prior to incarceration).⁵

Parenthood status

We utilize the variable “NCHILD” available via IPUMS to calculate the share of men living with children of their own among individuals who are not incarcerated. This variable is defined for every year.

Citizenship status

This variable is not available in 1880 and 1960. In 1870, 1900, and 1910, citizenship status was defined for foreign-born men older than 20. From 1920 onwards, it was defined for all foreign-born individuals. Individuals born in any US state are classified as citizens in all of these samples.

Additional Data Sources

ICE Facilities and Deportations Data

We identify Immigration and Customs Enforcement (ICE) detention facilities from the list provided in the 2022 ICE Detentions Statistics Appendix (U.S. Immigration and Customs Enforcement 2024). We identify two groups of facilities: The first group includes 18 ICE-owned

⁵ This assumption may not be true for those incarcerated for federal offenses because individuals might be sent to federal prisons outside of their state of residence. Nevertheless, the share of inmates in federal prisons is generally small (5–7% of incarcerated individuals in 1990 and 2000; Beck and Harrison 2001).

service processing centers and privately-owned contract detention facilities.⁶ In 2017, these types of facilities were 6% of the total number of facilities used for detention, but held approximately 28% of detainees.⁷ The second group extends this list to 107 facilities (for a total of 125 facilities) operated under agreements with local and state governments and federal agencies. This group includes facilities under intergovernmental service agreements and US Marshall’s administered facilities.

We geolocate these facilities and assign them to their corresponding PUMA in 1990 (1,726 total PUMAs), in 2006–2011 (2,069 total PUMAs), and 2012–2019 (2,351 total PUMAs) using shapefiles provided by IPUMS. For 1970 and 1980, we follow the same procedure using county group shapefiles provided by IPUMS (309 and 1,154 county groups in 1970 and 1980, respectively). Given changes in PUMA/county group geographic areas across time as well as the proximity of certain facilities to each other, we end up tagging 14–17 areas as including ICE facilities in the first group and 63–110 areas as including facilities in the second group. We exclude these areas from the sample in the second and third series of Figure A21, respectively.

To consider how the incarceration rate would change after including deportations, we use the 2006–2019 reports from the Department of Homeland Security on Immigration Enforcement Actions (U.S. Immigration and Customs Enforcement 2023). We focus on removals of individuals with criminal histories.

We also use data on annual removals from 1892 through today from U.S. Department of Homeland Security (2018).

Health

We use data from the General Social Survey (GSS) to measure health outcomes (Davern et al. 2023). We focus on the 1977–2021 period, in which individuals can be classified as foreign-born. We group annual data into five-year bins (e.g., the 2000 point includes the 1998–2002 survey waves). We rely on the “health” variable, identifying individuals who report an “excellent” or “good” health condition. Given small samples, we focus on men ages 18–65.

Admissions for Drug-Related Offenses

We use data from the National Corrections Reporting Program (NCRP; U.S. Department of Justice 2016) between 1991 and 2010. We derive the stock of incarcerated individuals for each year by keeping all records of individuals admitted to prison before or during that calendar year who are released after that same year. We then sum the number of drug-related incarcerations in each state and year and compute average drug incarceration rates at the state level for the 1991–1993 and 2008–2010. To calculate incarceration rates, we use state population counts from the 1990 and 2010 Census (Iowa State University 2024).

⁶ See also <https://www.ice.gov/doclib/foia/media-requests/09foia5638detentionfacilitylist.xls>.

⁷ “ICE Does Not Fully Use Contracting Tools to Hold Detention Facility Contractors Accountable for Failing to Meet Performance Standards.” Office of Inspector General, Department of Homeland Security. <https://www.oig.dhs.gov/sites/default/files/assets/2019-02/OIG-19-18-Jan19.pdf>.