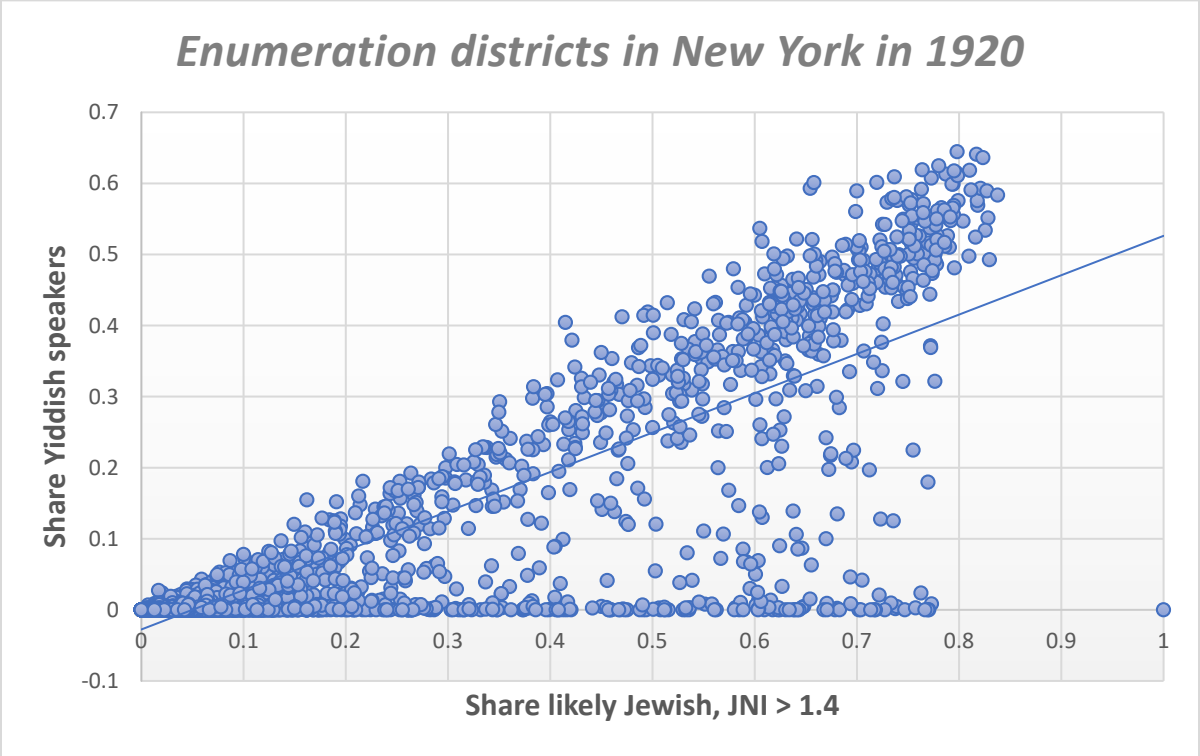
Leaving the Enclave: Historical evidence on immigrant mobility from the Industrial Removal Office

June 2023

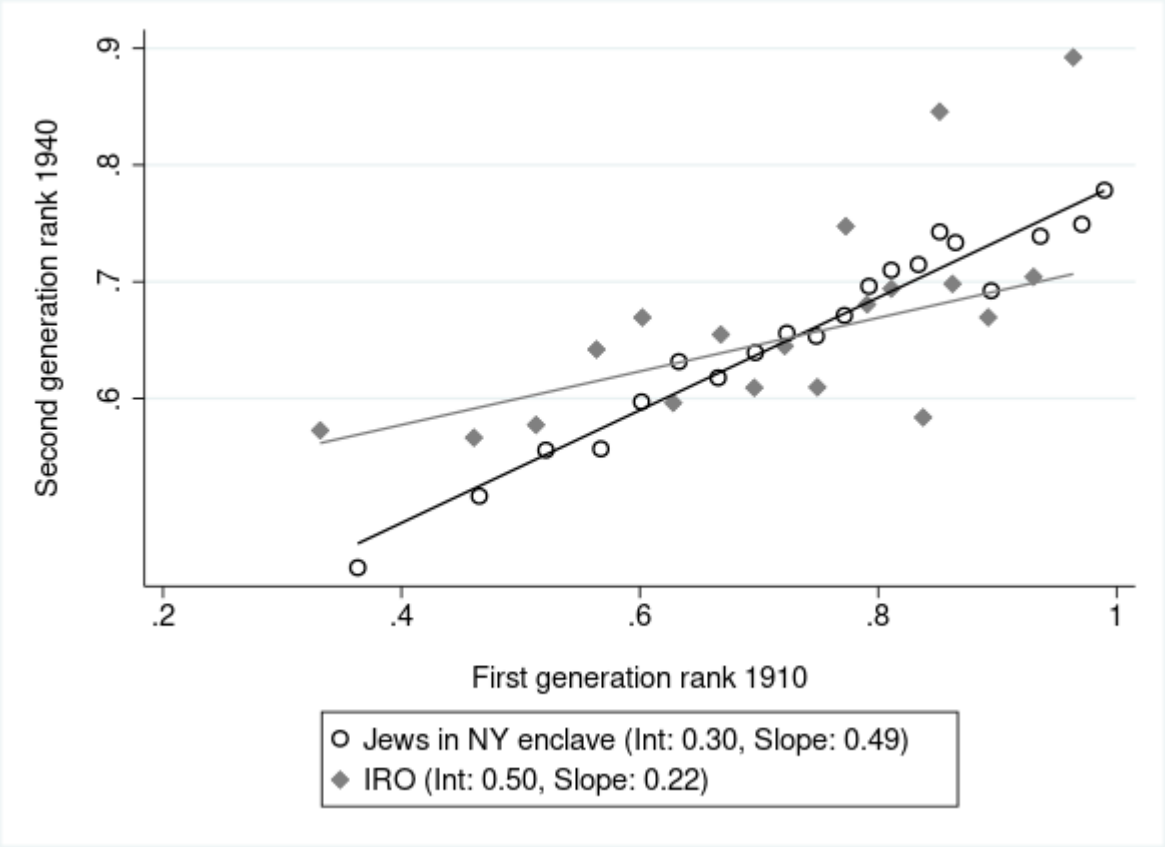
Appendix -- For On-line Publication Only

Appendix Figure 1. Scatterplot of New York enumeration districts in 1920, showing the share of Yiddish speakers and the share "likely Jewish"



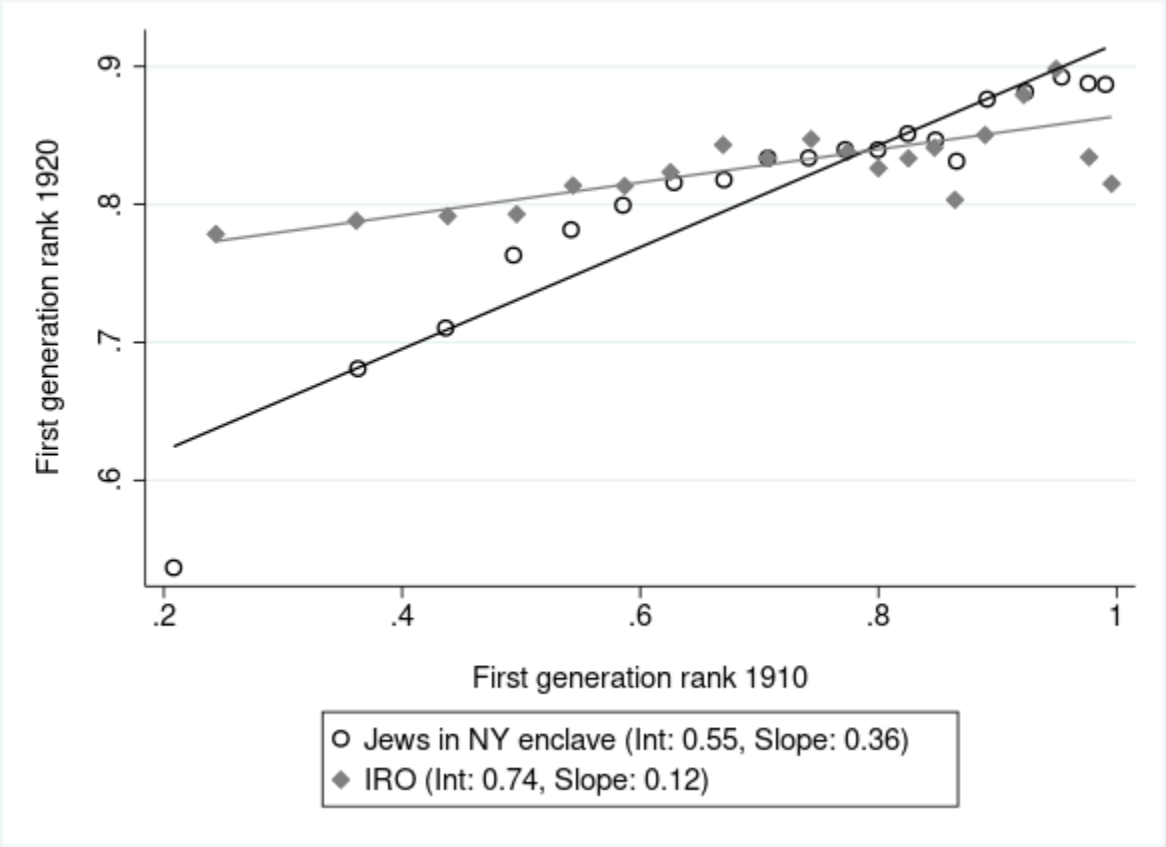
Notes: A figure showing the relationship between the share of Yiddish speakers and the share of likely Jews in New York enumeration districts in 1920. The correlation between these ED-level measures in New York City is +0.85.

Appendix Figure 2A: Rank-rank correlation for log income score of first-generation men in 1910 and second-generation sons in 1940



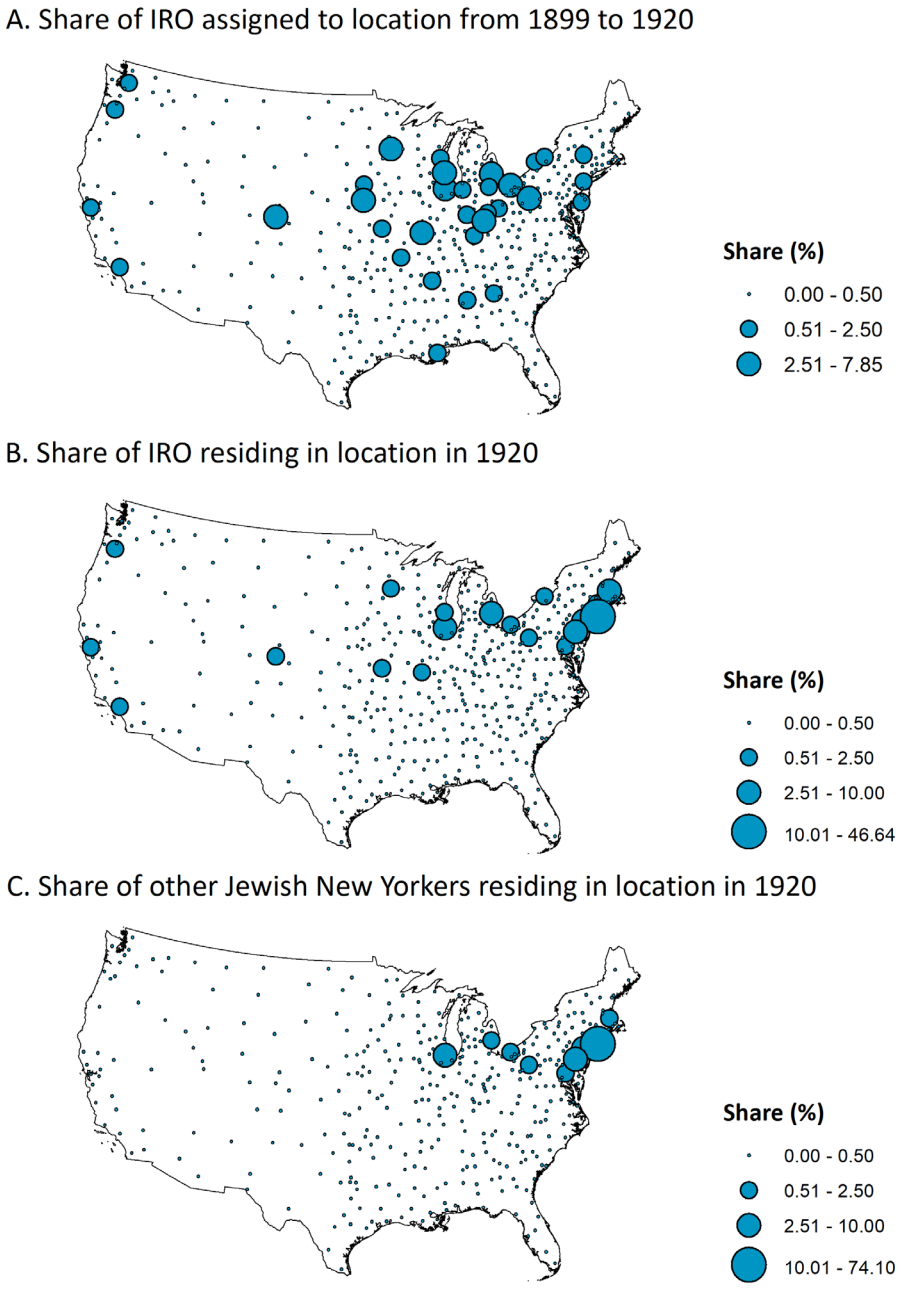
Notes: Binned scatterplot graphing the 1910 income rank of IRO participants and those of their sons in 1940 against the corresponding values for other Jewish immigrants in New York enclaves (circa 1910). The first and second generation in each group are assigned percentile ranks based on their log income score. The figure plots the mean income rank for each group as well as the corresponding regression lines.

Appendix Figure 2B: Rank-rank correlation for log income score of first-generation men in 1910 and in 1920



Notes: Binned scatterplot graphing the 1910- and 1920-income rank of IRO participants against other Jewish immigrants in New York enclaves (circa 1910). Men in each group are assigned percentile ranks based on their log income score. The figure plots the mean income rank for each group as well as the corresponding regression lines.

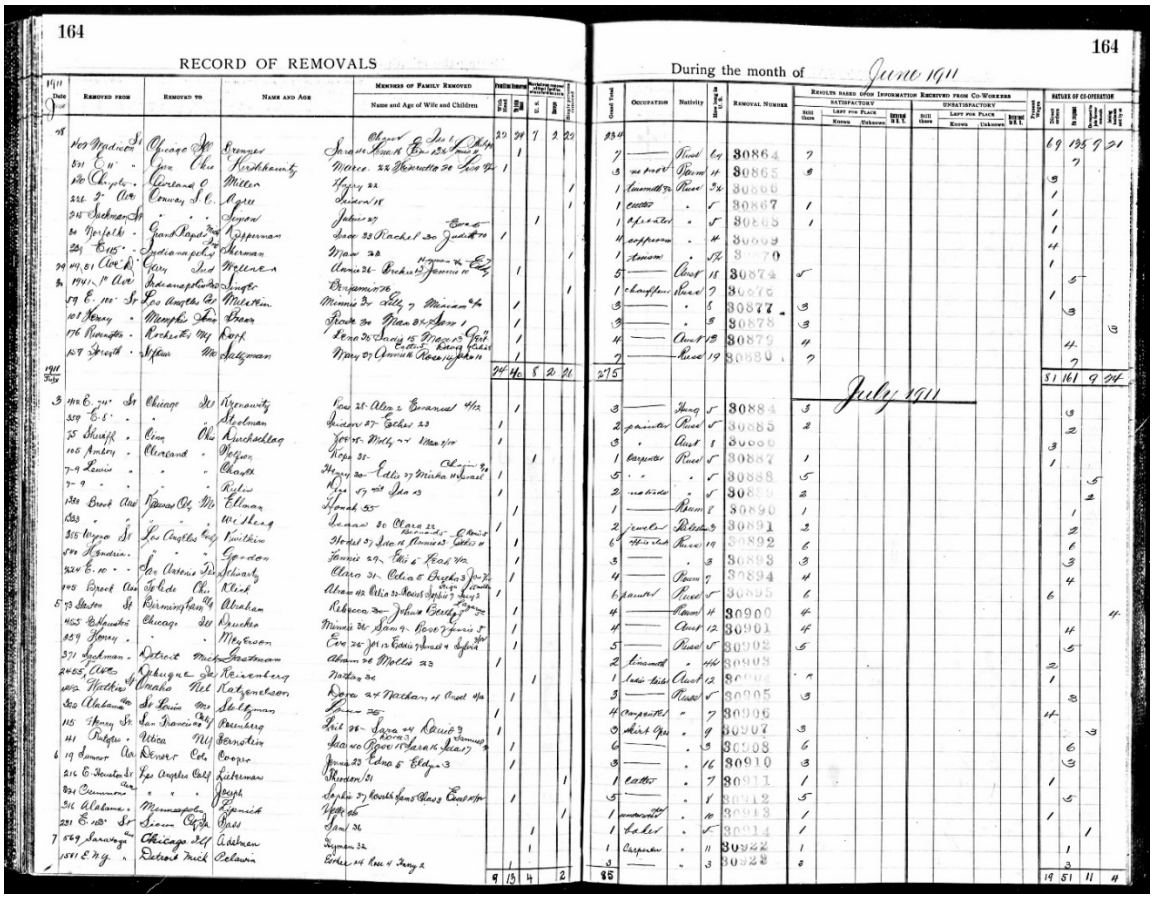
Appendix Figure 3: Distribution of IRO and other foreign-born Jewish households across state economic areas, 1899-1920



Notes: Panel A aggregates the placement cities reported by IRO to the scale of state economic areas (SEA) to display share (%) of IRO participants that were placed in different SEAs. Panel B uses the linked IRO-1920 Census sample to observe the 1920 post-resettlement locations of IRO

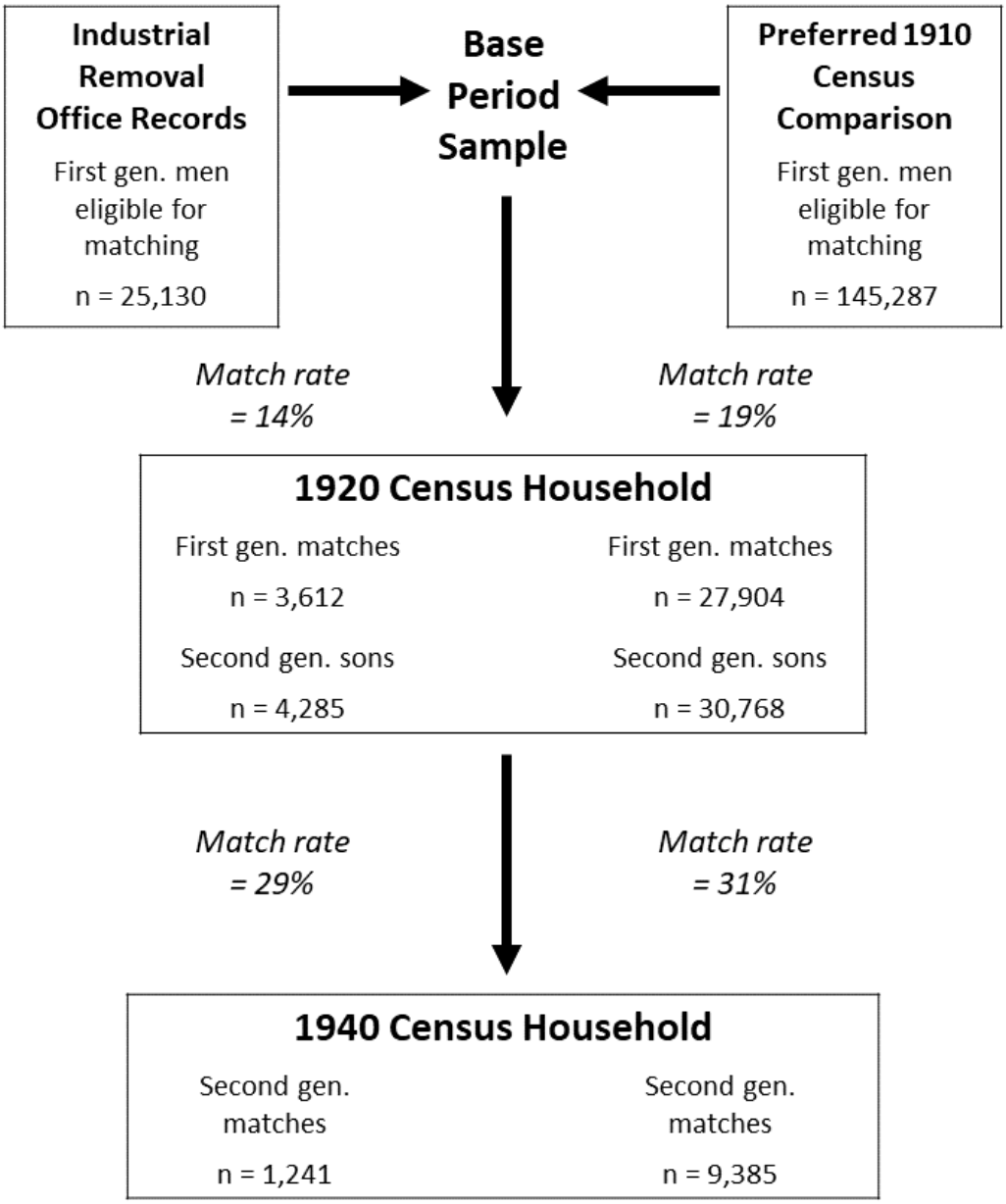
participants. Panel C maps the locations of non-IRO Jewish New Yorkers, our main comparison group, from the 1920 census.

Appendix Figure 4: Ledger page from the record books of the Industrial Removal Office



Notes: Photograph of a page from the original IRO ledgers held by the American Jewish Historical Society (New York) and made available online by Ancestry.com.

Appendix Figure 5: Matching procedure and observation counts for IRO and preferred comparison (resident in New York enclave in 1910)



Notes: The observations in this diagram reflect the larger underlying samples of interest. The sample sizes in our analyses may be smaller due to missing data or the analysis-specific sample restrictions discussed in table notes.

Appendix Table 1: Immigrant and Jewish enclaves in major US cities, summary statistics, 1910

	10 largest urban areas		**New York only**	
	Immigrant enclaves	Other neighborhoods	Jewish enclaves	Other neighborhoods
	(1)	(2)	(3)	(4)
Neighborhoods (N)	2,576	9,193	720	2,657
Total population	1,849	1,462	1,785	1,503
Immigrant share	0.53	0.23	0.57	0.33
Jewish share	0.18	0.04	0.42	0.06
English-speaking share	0.74	0.93	0.73	0.90
Mean income score, all (1940$)	772.64	772.71	812.40	883.30
Mean income score, Jewish (1940$)	814.05	819.51	829.12	939.66
White-collar share	0.21	0.31	0.28	0.33
Manufacturing share	0.34	0.24	0.36	0.19
Homeowner share	0.12	0.32	0.05	0.18

Notes: Characteristics of immigrant and Jewish enclaves in 1910. Columns 1 and 2 are based on the full population of enumeration districts in the 10 most populated state economic areas. For these columns, immigrant enclaves are defined as enumeration districts that are at least 40 percent foreign born. Boundaries of Jewish enclaves in New York are shown in Figure 8. We define New York from its state economic area boundaries.

Appendix Table 2: Classification of Jews in the 1920 Census by Jewish Names Index and Yiddish speaking

		Disagreement of Yiddish speaker and JNI		
		(1) Yiddish & JNI > 1.4 (agree)	(2) Yiddish & JNI < 1.4 (disagree)	(3) % Yiddish & JNI < 1.4 (disagree)
A	FB, male, age 26-59	297,072	80,579	21.34%
	Russian born	218,932	56,368	20.48%
	Other foreign born	78,140	24,211	23.65%
B	FB, male, age 26-59, NYC	146,350	32,391	18.12%
C	FB, male, age 26-59, NYC, enclave	70,837	13,427	15.93%
		Disagreement of non-Yiddish speaker and JNI		
		(4) Yiddish & JNI > 1.4 (agree)	(5) Non-Yiddish & JNI > 1.4 (disagree)	(6) % Non-Yiddish & JNI > 1.4 (disagree)
D	FB, male, age 26-59	297,072	200,012	40.24%
	Russian born	218,909	73,199	25.06%
	Other foreign born	78,140	126,813	61.87%
E	FB, male, age 26-59, NYC	146,350	69,639	32.24%
F	FB, male, age 26-59, NYC, enclave	70,837	13,660	16.17%

Notes: Columns 1 and 2 (4 and 5) contain counts of individuals in the 1920 census who report a Yiddish (non-Yiddish) mother tongue or with a Jewish Names Index greater than 1.4. We first limit the samples based on whether they are foreign born ("FB") males, aged between 26 and 59 in 1920 (A and D), and then restrict by whether they lived in the broader New York area (B and E) or specifically in a New Jewish enclave (C and F). The implied false negatives are calculated in Column 3 as: Column 2/(Column 1 + Column 2). The implied false positives are calculated in Column 6 as: Column 5/(Column 4 + Column 5).

Appendix Table 3: Evaluation of the sensitivity of Jewish classification to 1.4 threshold, based on Yiddish speakers in the 1920 Census

	(1)	(2)	(3)	(4)	(5)
	Yiddish speaker & Jewish Index > threshold	Yiddish Speaker & Jewish Index < threshold	Non-Yiddish speaker & Jewish index > threshold	% Yiddish Speaker & Jewish Index < 1.4 (possible false negative)	% Non-Yiddish speaker & Jewish index > 1.4 (possible false positive)
Jewish index > 1.0	80,136	4,128	16,398	4.9%	16.99%
Jewish index > 1.2	76,744	7,490	15,166	8.89%	16.50%
Jewish index > 1.4	70,837	13,427	13,660	15.93%	16.17%
Jewish index > 1.6	64,753	19,511	12,185	23.15%	15.84%
Jewish index > 1.8	51,024	33,240	9,296	39.45%	15.41%

Notes: A table that evaluates the sensitivity of our Jewish classification using the mother tongue variable (Yiddish speakers) in the 1920 census. We restrict the observation foreign born males, aged 26-59, who lived in a Jewish enclave in New York City in 1920. The possible false negative rate is calculated as: Column 2/(Column 1 + Colum 2). The possible false positive rate is calculated as: Column 3/(Column 3 + Colum 1).

Appendix Table 4: Log income score and total years of schooling for second-generation sons of IRO participants in 1940

	Outcome = Years of schooling Cross-section	
	1940	1940
Second generation		
IRO	-0.106	0.00325
	(0.128)	(0.139)
N	7723	7723
Controls		
Birth cohort	Y	Y
Arrival Year	Y	Y
Russian birthplace	Y	Y
~1910 Occ.	N	Y
~1910 Inc. rank	N	Y

Robust standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01

Notes: Schooling differences in 1940 between sons of IRO and other Jewish immigrants living in New York enclaves circa 1910. Reference category are the sons of Jews whose fathers lived in New York enclaves in 1910. The second-generation were aged 18 to 41 in 1940 and are the sons of immigrants with a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. The difference-in-difference coefficients (Columns 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1940 for second generation). Controls in the diff-in-diff models are estimated with a main effect and an interaction with the period dummy. Standard errors clustered at household for second generation.

Appendix Table 5: Difference in difference estimates for the log income score change for IRO participants by program exposure, with additional controls and sample restrictions

	A. Name-based controls		B. Direct removals only	C. Arrived >2 years before IRO
	(1)	(2)	(3)	(4)
A. Years of treatment				
IRO: 14-20 years (early)	0.154***	0.120***	0.146***	0.173***
	(0.023)	(0.016)	(0.021)	(0.019)
IRO: 8-13 years (middle)	0.0640***	0.0664***	0.0532***	0.0302*
	(0.013)	(0.014)	(0.014)	(0.013)
IRO: 1-7 years (late)	-0.0603***	-0.0591***	-0.0963***	-0.0989***
	(0.014)	(0.015)	(0.015)	(0.014)
N	44070	44070	42176	42404
B. Compliance with relocation				
IRO: Returned to NYC	0.0609***	0.0527***	0.0202	0.0275*
	(0.014)	(0.012)	(0.013)	(0.012)
IRO: Stayed outside NYC	0.0409**	0.0344**	0.00784	-0.00675
	(0.015)	(0.013)	(0.014)	(0.013)
N	44070	44070	42176	42404
Controls				
Birth cohort	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y
~1910 Occ.	Y	Y	Y	Y
~1910 Inc. rank	Y	Y	Y	Y
Name-based network	Y	N	N	N
Last name	N	Y	N	N

Robust standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01

Notes: IRO program exposure and log income score changes by 1920 with control for frequency of last name in immigrant population in New York City in 1910 (C1), last name fixed effects (C2), direct removals only (C3), and participants who were in the US two years prior to removal (C4). Reference category are Jews living in New York enclaves in 1910. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920.

Appendix Table 6: Cultural assimilation of IRO participants in 1920 by program exposure

	Own Jewish index	English speaking HH	Wife's Jewish index	Child's Jewish index
	(1)	(2)	(3)	(4)
	~1910	1920	1920	1920
A. Years of treatment				
IRO: 14-20 years (early)	0.00986	0.0397***	-0.00568	0.00566
	(0.008)	(0.012)	(0.013)	(0.013)
IRO: 8-13 years (middle)	0.0136*	0.0201	-0.0241*	0.0138
	(0.007)	(0.012)	(0.014)	(0.011)
IRO: 1-7 years (late)	0.000998	0.00629	-0.0217*	0.0310***
	(0.008)	(0.012)	(0.012)	(0.011)
N	6883	6883	6883	12300
Controls				
Birth cohort	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y
Own Jewish index	N	N	Y	Y
Child: age, sex, foreign born	N	N	N	Y

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Cultural assimilation by IRO program exposure as measured by own Jewish name index in base period, wife's Jewish name index in 1920 and child's Jewish name index in 1920. Reference category are Jews living in New York enclaves in 1910. The first-generation sample is restricted to household heads in 1920 who were not co-resident with a spouse in the base period, and with a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). The regression underlying Column 4 is estimated at the child level, rather than the father level. The sample includes children between the ages of zero and 10 who were observed in 1920 households.

Appendix Table 7: Log income score of IRO participants in 1920 and second-generation sons in 1940 [replication of Table 5 with ABE Conservative]

	Cross-section		Diff-in-diff	
	(1)	(2)	(3)	(4)
	~1910	1920	~1910-1920	~1910-1920
A. First generation				
IRO	-0.181***	-0.0514***	-0.208***	-0.0625***
	(0.010)	(0.012)	(0.011)	(0.005)
In 1920			0.935***	1.511***
			(0.031)	(0.050)
IRO *x* In 1920			0.184***	0.0482***
			(0.016)	(0.014)
N	12939	12939	25878	25878
	~1910	1940	~1910-1940	~1910-1940
B. Second generation				
IRO	-0.207***	-0.0178	-0.207***	-0.0468***
	(0.023)	(0.021)	(0.023)	(0.010)
In 1940			1.223***	1.654***
			(0.076)	(0.136)
IRO *x* In 1940			0.189***	0.0416
			(0.031)	(0.024)
N	2645	2645	5290	5290
Controls				
Birth cohort	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y
~1910 Occ.	N	N	N	N
~1910 Inc. rank	N	N	N	Y

Robust standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Log income score difference between IRO and other Jews living in New York enclaves in 1910. Reference category are Jews living in New York enclaves in 1910. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. The difference-in-difference coefficients (Column 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation, post-1940 for second generation). Controls in the diff-in-diff models are estimated with a main effect and an interaction with the period dummy. Linear term for age at first observation included as additional continual control variable for IRO. For the

second-generation sons, aged 18 to 41 in 1940, the dependent variable is the log of actual income in 1940 dollars). Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). Standard errors clustered at household for second generation.

Appendix Table 8. Other economic outcomes for IRO participants, 1920 [replication of Table 6 with ABE Conservative]

	In labor force	Employer	Self-employed	Professional worker	Manufact worker	Citizen	Owns home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1920	1920	1920	1920	1920	1920	1920
IRO	-0.0024	-0.0099	0.0382*	0.0216	-0.0338*	-0.0049	0.0229
	(0.002)	(0.010)	(0.016)	(0.017)	(0.017)	(0.017)	(0.015)
N	12939	12939	12939	12939	12939	12939	12939
Mean of dependent var, comparison group	0.99	0.12	0.29	0.34	0.27	0.58	0.15
Controls							
Birth cohort	Y	Y	Y	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y	Y	Y	Y
Russian	Y	Y	Y	Y	Y	Y	Y
~1910 Inc score	Y	Y	Y	Y	Y	Y	Y

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Other economic differences in 1920 between IRO and other Jews living in New York enclaves in 1910. Reference category are Jews living in New York enclaves in 1910. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. All outcomes are derived from the following IPUMS variables Column 1 (LABFORCE), Columns 2-3 (CLASSWKR), Column 4 (IND1950), Column 5 (OCC1950), Column 6 (CITIZEN), Column 7 (HOMEOWNER). Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). For reference, the table includes the mean of the dependent variable for the comparison group.

Appendix Table 9: Cultural assimilation of IRO participants, 1920 [replication of Table 7 with ABE Conservative]

	Own Jewish index	Speaks English	Wife's Jewish index	Jewish index of children		
				All	*Sons*	*Daughters*
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	1920	1920	1920	1920
IRO	0.00462	0.00955	-0.0385**	0.0272*	0.0540**	0.00892
	(0.008)	(0.014)	(0.013)	(0.012)	(0.018)	(0.017)
N	3660	3660	3660	6522	3352	3170
Controls						
Birth cohort	Y	Y	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y	Y	Y
Own Jewish index	N	Y	Y	Y	Y	Y
English speaking HH	Y	Y	Y	N	N	N
Child: age, sex, foreign	N	N	N	Y	Y	Y
Household clustered SEs	N	N	N	Y	Y	Y

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Cultural assimilation differences as measured by own Jewish name index in base period, wife's Jewish name index in 1920 and child's Jewish name index in 1920. Reference category are Jews living in New York enclaves in 1910. The first-generation sample is restricted to household heads in 1920 who were not co-resident with a spouse in the base period, and with a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). The regression underlying Column 3 is estimated at the child level, rather than the father level. The sample includes children between the ages of zero and 10 who were observed in 1920 households.

Appendix Table 10. Log income score of IRO participants in 1920, by refugee status [replication of Table 8 with ABE Conservative]

	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
Reference = Not IRO, not refugee			
IRO, not refugee	-0.172***	-0.0341	0.0575*
	(0.017)	(0.025)	(0.026)
IRO, refugee	-0.198***	-0.0627**	0.0281
	(0.018)	(0.019)	(0.022)
Not IRO, refugee	0.00364	0.0132	0.00239
	(0.012)	(0.016)	(0.016)
N	12939	12939	25878
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	N
~1910 Inc. rank	N	N	Y

Robust standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: IRO income score change by 1920, differentiated by refugee status. We define refugees as immigrants who left Russia between 1903 and 1906, a period of widespread pogroms in Russia and Eastern Europe. In total, 1,262 (27%) IRO participants and 19,726 (17%) members of the comparison group are classified as refugees. Observations are restricted to males with a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix).

Appendix Table 11: Log income score of IRO participants in 1920 by program exposure [replication of Table 9 with ABE Conservative]

	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
A. Years of treatment			
IRO: 14-20 years (early)	-0.180***	-0.0276	0.168***
	(0.045)	(0.042)	(0.024)
IRO: 8-13 years (middle)	-0.214***	-0.0408*	0.0721***
	(0.018)	(0.019)	(0.020)
IRO: 1-7 years (late)	-0.158***	-0.0692**	-0.0447*
	(0.026)	(0.026)	(0.018)
N	12939	12939	25878
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	N
~1910 Inc. rank	N	N	Y

Robust standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: IRO program exposure and log income score changes by 1920. Reference category are Jews living in New York enclaves in 1910. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. The difference-in-difference coefficients (Column 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation). Controls in the diff-in-diff models are estimated with a main effect and an interaction with the period dummy. Linear term for age at first observation included as additional continual control variable for IRO. Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix).

Appendix Table 12: Economic and cultural assimilation of IRO participants by return to New York [replication of Table 10 with ABE Conservative]

	Income score (cross-section)		Income score (diff-in-diff)	
	(1)	(2)	(3)	(4)
	~1910	1920	~1910-1920	~1910-1920
A. Compliance with relocation				
IRO: Returned to NYC	-0.168***	-0.0199	0.176***	0.0684***
	(0.015)	(0.015)	(0.022)	(0.017)
IRO: Stayed outside NYC	-0.193***	-0.0782***	0.191***	0.0313
	(0.013)	(0.015)	(0.021)	(0.017)
N	12939	12939	25878	25878

	Own Jewish index	English speaking HH	Wife's Jewish index	Child's Jewish index
	~1910	1920	1920	1920
B. Compliance with relocation				
IRO: Returned to NYC	0.0105	-0.0212	0.0259	-0.0123
	(0.010)	(0.017)	(0.033)	(0.020)
IRO: Stayed outside NYC	-0.000671	-0.0540**	-0.0578	0.0292*
	(0.010)	(0.018)	(0.033)	(0.014)
N				
Controls				
Cohort, arrival year, birthplace	Y	Y	Y	Y
~1910 Occ [a]	N	N	N	Y
~1910 Inc. rank [a]	N	N	N	Y
Own Jewish index [b]	N	N	Y	Y
Child: age, sex, foreign born [b]	N	N	N	Y

Robust standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Economic and cultural assimilation by participants decision to return to New York. Reference category are Jews living in New York enclaves in 1910. The first-generation sample is restricted to household heads in 1920 who were not co-resident with a spouse in the base period, and with a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). The regression underlying Column 3 is estimated at the child level, rather than the father level. The sample includes children between the ages of zero and 10 who were observed in 1920 households. The superscripts refer to controls that are used only in the models with income score outcomes ([a]) and for the cultural outcomes only ([b]).

Data Appendix

DA1. Record linkage & sample construction

Our record linkage approach is based on the methods originally developed by Ferrie (1996) and further refined by Abramitzky, Boustan and Eriksson, 2012, 2014) ("ABE"). There are now several reviews of these original approaches and their recent extensions (Abramitzky et al., 2019; Bailey et al., 2017; Feigenbaum, 2016; Ruggles et al., 2017). These matching approaches link individuals across data sources by their first and last name, birthplace and year of birth (inferred from age) with the assumption that these characteristics are stable across data sources.

Because age, and consequently year of birth, may be misreported or contain transcription errors, our matching algorithms take an iterative approach. We first attempt to link individuals across data sources based on having an identical name, birthplace and year of birth in the two data sources. If we fail to find such an individual, we allow for measurement error in year of birth by up to one year, and beyond that, up to two years. For example, if we were attempting to link a 16-year-old from the 1910 Census to the 1920 Census, we would first search for an individual with an identical name and birthplace who was born in 1894. If no match could be found, we would then widen our search to include individuals with a year of birth of 1893 and 1895, and beyond that 1892 and 1896. If we find, in any of these steps, more than one individual with matching characteristics, we abandon the search for this individual and exclude the individuals from the sample.

Although we rely on one main linkage approach to construct our main sample ("ABE EXACT NAME"), we test the robustness and sensitivity of our results by linking our sample using a *more* conservative algorithm ("ABE CONSERVATIVE") and a *less* conservative algorithm ("ABE NYSIIS"):

- **ABE EXACT NAME:** Individuals are linked across data sources based on having an identical first name, last name, year of birth and birthplace. If we fail to find an individual with exact matching characteristics, we follow the iterative year of birth sequence described above. We undertake moderate name cleaning for unusual characters and common name transitions. For example, we do not distinguish between the names Abe and Abraham or Joe and Joseph.
- **ABE NYSIIS:** Individuals are linked across data sources following the criteria described for ABE EXACT NAME but we undertake additional name cleaning. Specifically, we implement the phonetic coding of the New York State Identification and Intelligence System ("NYSIIS"). This coding system adjusts for a wide range of misspelling and name changes by phoneticizing the names recorded in our written data sources. While the NYSIIS approach improves the linkage rate, it tends to increase the rate of false-positive matches (Bailey et al., 2017). Thus, we consider this as our least conservative linkage approach.
- **ABE CONSERVATIVE:** Individuals are linked across data sources following the criteria described for ABE EXACT NAME but impose a higher uniqueness threshold for acceptable linkages. Specifically, we undertake an initial screen on our data so that we only attempt to link individuals who are unique in terms of name and birthplace within two years of their year of birth. Worded differently, for each individual we attempt to link between the 1910 Census and 1920 Census, we screen the sample to only include cases where there are no other individuals with the same name and birthplace born within two years. For example, if an individual was born in 1894, they are only eligible for matching if there are no individuals with the same name and birthplace born between 1892 and 1886.

As mentioned above, our main analyses rely on the ABE EXACT NAME linkage approach and we provide alternate analyses for our main results using the ABE NYSIIS and ABE CONSERVATIVE samples.

We applied these linkage algorithms to four main data sources: The Industrial Removal Office records, and the 1910, 1920 and 1940 decennial censuses of the United States. As it is possible for an individual to be enumerated in both the IRO records and the 1910 decennial census, this complicated the construction of our baseline sample. Specifically, we needed to pre-screen the 1910 Census to remove any individual already present in the IRO records. We did this by searching for individuals with identically matching names, birthplaces and years of birth.

We linked our data as follows:

1. Merge the IRO records and the 1910 census records. In the merged data:
    a. If the IRO individual is not duplicated in the census, the IRO case is eligible for linkage.
    b. If an individual is duplicated only once in the merged dataset (found in the 1910 census and IRO), assume that that this is the same individual and drop the census duplicate.
    c. If an individual is duplicated more than once in the census, the IRO and the census cases are ineligible for linkage and are removed from the sample.
2. Use the ABE algorithms above to link individuals from the merged baseline dataset to the 1920 census records.
3. From this linked dataset, we then also search for second-generation sons in the 1920 household. Again, using the ABE methods above, we can then link second-generation sons from the 1920 census to the 1940 census.

In Data Appendix Table 1, we document the level of sample attrition throughout the linkage procedure. For the baseline to 1920 link, we also document the specific linkage rates for our alternate matched samples in Data Appendix Table 2.

Data Appendix Table 1. Sample attrition from primary data sources

	IRO	Lived in NYC enclave, 1910	Lived outside NYC enclave, 1910	Lived outside NYC, 1910
Baseline to 1920 match				
Foreign-born, Jewish males, aged 16-49 at baseline	25,130*	145,287	45,226	228,565
Does not share characteristics with other individuals in census (uniqueness screen)	21,547	117,796	38,356	191,994
Matched with BASIC procedure (Match rate % to 1920 Census)	3,612 (14%)	27,904 (19%)	10,039 (22%)	42,971 (19%)
Valid occupation, income score, locatable neighborhood	2,352	19,761	7,000	31,109
1920 to 1940 match				
Sons aged under 21 in 1920 household	4,285	30,768	10,090	46,752
Matched with BASIC procedure (Match rate % to 1940 Census)	1,241 (29%)	9,385 (31%)	3,372 (33%)	16,056 (34%)

Notes: The original IRO records contain 39,004 participants. We lose approximately 10,000 cases from the original population due to being female or having incomplete information on name, age or birthplace. The remaining attrition to 25,130 is due to individuals being outside of the 16-49 age, not having a sufficiently or being native-born.

Data Appendix Table 2. Record linkage rates across samples by linkage procedure

	IRO	Lived in NYC enclave, 1910	Lived outside NYC enclave, 1910	Lived outside NYC, 1910
	N	**N**	**N**	**N**
Does not share characteristics with other individuals in census (uniqueness screen)	21,547	117,796	38,356	191,994
Matched with BASIC ABE procedure (Match rate % to 1920 Census)	3,612 (14%)	27,904 (19%)	10,039 (22%)	42,971 (19%)
Matched with NYSIIS ABE procedure (Match rate % to 1920 Census)	5,064 (24%)	35,193 (30%)	12,276 (32%)	53,631 (28%)

Data Appendix Table 3. Comparison of full IRO records to linked IRO records

	(1) IRO records, non-matched (mean)	(2) IRO records, matched (mean)	(3) Difference (standard error)
Age in 1910	30.34	28.46	1.87 (0.16)
Year removed	1907.94	1908.82	-0.88 (0.071)
Jewish index	1.42	1.75	-0.33 (0.01)
Moved with wife	0.15	0.19	-0.03 (0.01)
Log income score in 1910	6.53	6.52	0.015 (0.008)
Lived in New York enclave	0.68	0.67	0.01 (0.01)
N	31,099	3,795	

Data Appendix Table 4. Comparison of 1910 Census eligible records and linked sample

	(1) 1910 Census, non-matched (mean)	(2) 1910 Census, matched (mean)	(3) Difference (standard error)
Age in 1910	31.36	31.52	-0.16 (0.035)
Jewish index	1.77	1.78	-0.01 (0.01)
Log income score in 1910	6.79	6.83	-0.42 (0.002)
Lived in New York enclave	0.35	0.35	0.01 (0.01)
N	326,336	78,290	

DA2. Weighting

Data Appendix Table 2 shows that our linkage rates vary from 14-22% depending on our baseline sample. These discrepancies in the linkage rate partly reflect differences in the attributes of the baseline samples such as year of birth or the distinctiveness of names. Thus, we construct sampling weights based on a set of these baseline characteristics, which we use to ensure that these linkage biases are not distorting our main results. Data Appendix Table 5 presents univariate estimates of how baseline (1910 census/IRO) characteristics relate to the probability of successful linkage from the baseline data to the 1920 Census.

The unweighted estimates provide an assessment of general linkage bias. In terms of baseline characteristics, IRO participants are generally less likely to be linked than the average Jewish male in the 1910 census. Age and the Jewish index are also positively correlated with linkage. Based on these characteristics, we use a probit regression to construct a set of sampling weights to rebalance our sample. The weighted estimates show that when we apply these weights variable-by-variable, most of the linkage bias associated with these characteristics disappears.

Data Appendix Table 5. Unweighted and weighted estimates of the probability of being linked from 1910/IRO to the 1920 census

	Outcome = Successfully linked from base period *to 1920 Census*			
	Univariate estimate (**unweighted**)		Univariate estimate (**weighted**)	
	Intercept	*Slope*	*Intercept*	*Slope*
IRO	0.192***	-0.0340***	0.191***	-0.00172
	(0.000598)	(0.00283)	(0.00144)	(0.00334)
Age	0.190***	2.31e-05	0.187***	0.000119
	(0.00210)	(6.34e-05)	(0.00483)	(0.000167)
Age squared	0.191***	-1.66e-07	0.189***	1.96e-06
	(0.00120)	(9.55e-07)	(0.00252)	(2.58e-06)
Jewish index	0.138***	0.0295***	0.227***	-0.0203
	(0.00622)	(0.00349)	(0.0203)	(0.0108)
Jewish index squared	0.165***	0.00793***	0.209***	-0.00584
	(0.00327)	(0.00101)	(0.0110)	(0.00310)
Birthplace				
Germany	0.188***	0.0229***	0.191***	0.000427
	(0.000613)	(0.00206)	(0.00150)	(0.00242)
Russia	0.190***	0.000435	0.191***	-0.000745
	(0.000932)	(0.00120)	(0.00329)	(0.00338)
Italy	0.192***	-0.114***	0.191***	0.00355
	(0.000588)	(0.00545)	(0.00139)	(0.00819)
Austria	0.192***	-0.00742***	0.191***	-0.00153
	(0.000630)	(0.00170)	(0.00158)	(0.00225)
N = 450,627				

* p < 0.05, ** p < 0.01, *** p < 0.001, Robust standard errors in parentheses

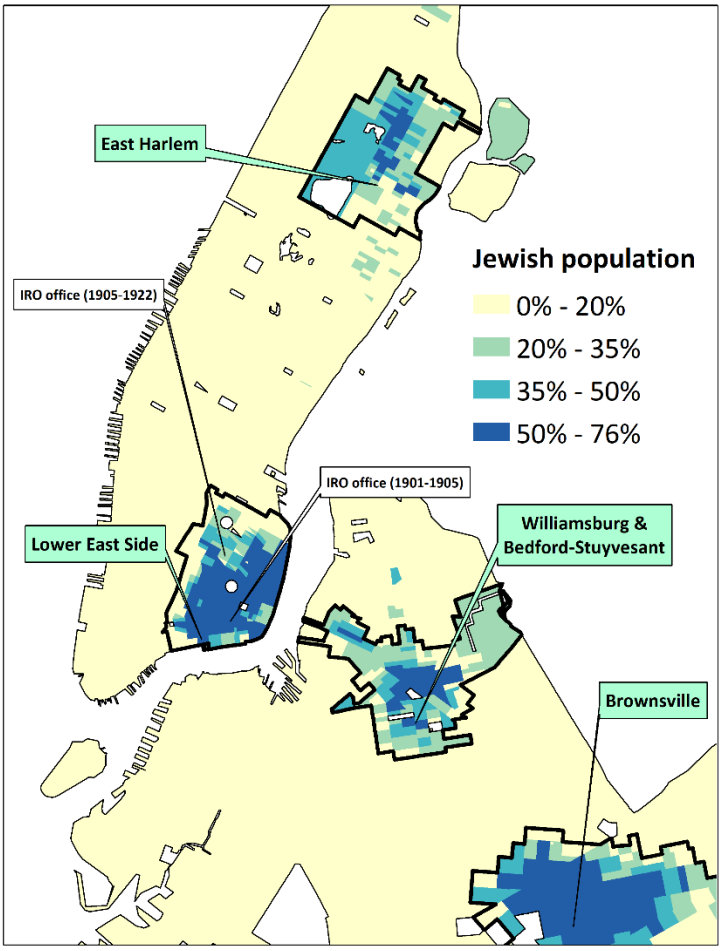
Notes: Sample are foreign-born with a Jewish index > 1.4. Each coefficient is derived from a univariate regression with one independent variable. We suppress the coefficients for small birthplaces categories (e.g. Mexico, Canada, Ireland, England).

DA3. Address matching procedure

To construct our comparison groups and examine neighborhood change for IRO participants, we needed to classify the neighborhoods of Jewish households throughout the early twentieth century. As we could rely on the census-reported enumeration district, this was generally straightforward for the comparison households. We then also describe the comparison households' neighborhoods in terms of their Jewish population characteristics, economic status and homeownership.

In addition to these quantitative attributes, we also classified Jewish households in New York in 1910 by whether they lived in a Jewish enclave. Using 1910 enumeration district .shp files provided by Allison Shertzer, we plotted the Jewish population share of enumeration districts (calculated from our Jewish names index). From Data Appendix Figure 1, we delineated four identifiable Jewish enclaves: Lower East Side; East Harlem; Bedford-Stuyvesant/Williamsburg; Brownsville. We manually delineated these neighborhoods through visual inspection of the Jewish population shares around known Jewish enclave areas. As the IRO primarily focused on moving or deflecting Jewish households away from Jewish neighborhoods, we segmented the New York comparison group by whether they lived in or outside of one of these four enclaves.

Data Appendix Figure 1. Jewish share of enumeration districts in New York in 1910 and neighborhood classification



Classifying the neighborhoods of IRO participants was more challenging. Although the IRO ledgers report a New York street address, these geographic identifiers are not easily located in space or linked to the 1910 census enumeration districts. Although contemporary street addresses can be located using a geocoder with relatively high levels of precision, this is significantly more challenging for historical addresses. This is due to significant increases in geolocation error resulting from historical changes in the numbering systems, street names, and road networks of American streets. It is thus unsurprising that efforts to locate historical addresses using contemporary geocoders yield error rates of at least 30-40% (Connor et al., 2019). Given that we focus on addresses recorded more than a century ago, we suspect that this error rate would be even higher.

With these issues in mind, we devised a new strategy for spatially locating IRO households. Instead of attempting to precisely situate IRO addresses in space, we instead focus on assigning these households to a 1910 enumeration district from the 1910 Census. By doing so, we could rely on the same approach that we employ for characterizing the neighborhoods of the census comparison group. We do this by using string and numeric matching approaches to pair the addresses reported in the IRO records with those reported in the 1910 Census. Effectively, we pair the IRO addresses with the same or similar addresses in the 1910 Census and use this link to impute the IRO records with a 1910 enumeration district. We do this in the following steps:

1. We clean the IRO and census street addresses of spaces and unusual characters. We then use the "MATCHIT" module in Stata to compare all street names in the IRO records to all New York street names in the 1910 Census. This procedure produces a pairwise score of street name similarity, where 1 represents an identically matching street name and 0 means there is no similarity at all. After this first stage, we restrict the IRO-Census addresses to the pair of streets with the highest similarity score (most similar words/fewest letter replacements).
2. Focusing on these most closely matching streets, we then calculate the difference between the IRO street number and all street numbers reported for the candidate street in the 1910 Census. From this calculation, we then limit the candidate street address to the one with the shortest numeric difference between the IRO record and the 1910 Census address. Thus, for every IRO street address, this leaves us with a single most likely matching address based on street name and street number (see example in Data Table 9).
3. From these two steps, we can link every IRO record to a likely matching address in the 1910 Census. We use this link to extract an enumeration district number from which we can measure the baseline neighborhood attributes of IRO households, and whether or not the IRO households were living in a Jewish enclave in New York.

One of the major advantages of our approach to using a geocoder is that we have two measures of uncertainty: the score based on the similarity between the IRO-reported street name and the street we linked it to in the 1910 Census, and the difference between the IRO and Census street numbers. Using these two error measures, we define a threshold for a *good address match*. We define a *good address match* as one where the similarity score is greater than 0.7 and differences in street numbers is less than 250. We use the *good address match* cases for our main analyses.

We based our decision to restrict to street numbers with a gap of less than 250 on the trends evident in the data. In Data Appendix Table 7, we show the Jewish share of enumeration districts based on the distance between possible street numbers. As we know that the activities of the IRO were generally focused on households living in Jewish neighborhoods, we use the Jewish share of the imputed enumeration district to inspect sensitivity to street number error. When the Census street number is less than 100 away from the IRO street number, the Jewish share of the ED ranges from 0.37-0.48. When we incorporate street numbers 100-249 away from the reported street number, the Jewish share drops to 0.30. This likely reflects the increased probability of street mismatching among these cases. However, it is not until we expand the street number error to 250-499 that we see a very substantial reduction in the Jewish share to 0.18. Consistent with our hypothesis, when the street number error is larger, the expected Jewish share declines sharply.

Data Appendix Table 6. Example streets for address matching procedure

IRO record	IRO address	Most closely matching street names	
		Name	*Similarity score*
Samuel Feldstein	224, DELANCEY	DELANCEY	1
		DELANCEYPLACE	0.88
		DELANEYST	0.69
		DELCAR	0.62
		GLANCEST	0.57
Jake Bergman	192, DELAUCEY	DELANCEY	0.71
		DELANCEYPLACE	0.66
		DELACER	0.62
		DELACEYSTREET	0.54
		DELANEYST	0.54

Data Appendix Table 7. Quality of street address matching and Jewish share of enumeration district

Difference between census street number IRO street number	Jewish share ED 1910	IRO cases
0-49 (small distance)	0.48	1719
50-99	0.37	310
100-249	0.30	272
250-499	0.18	119
500+ (large distance)	0.11	62

Street name similarity	Jewish share ED 1910	IRO cases
0.2-0.4 (not similar)	0.02	2
0.4-0.6	0.19	57
0.6-0.8	0.33	191
0.8-0.9	0.33	193
0.9-1 (identical)	0.43	2298