LEAVING THE ENCLAVE:
HISTORICAL EVIDENCE ON IMMIGRANT MOBILITY
FROM THE INDUSTRIAL REMOVAL OFFICE

Ran Abramitzky
Leah Platt Boustan
Dylan Connor

ABSTRACT

We study a program that funded 39,000 Jewish households in New York City to leave enclave neighborhoods circa 1910. Compared to their neighbors with the same occupation and income score at baseline, program participants earned 4 percent more ten years after removal, and these gains persisted to the next generation. Men who left enclaves also married spouses with less Jewish names, but they did not choose less Jewish names for their children. Gains were largest for men who spent more years outside of an enclave. Our results suggest that leaving ethnic neighborhoods could facilitate economic advancement and assimilation into the broader society, but might make it more difficult to retain cultural identity.

Ran Abramitzky
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and NBER
ranabr@stanford.edu

Dylan Connor
Arizona State University
dyl1289@nber.org

Leah Platt Boustan
Princeton University
Industrial Relations Section
Louis A. Simpson International Bldg.
Princeton, NJ 08544
and NBER
lboustan@princeton.edu

1. Introduction

During the Age of Mass Migration (1850-1913), many immigrants to the United States lived in immigrant neighborhoods and relied on ethnic networks.[1] At the time, both pro- and anti-immigration voices expressed concerns about poor conditions in immigrant neighborhoods, and how the isolation of immigrant enclaves might impede assimilation.[2] This paper studies the economic and cultural assimilation of one immigrant group that moved to the US in the early twentieth century – Eastern European Jews – and asks whether leaving enclave neighborhoods generated upward mobility. To do so, we leverage a unique program called the Industrial Removal Office (IRO) that financed 39,000 Jewish immigrant households to move out of Jewish enclaves in New York City between 1900-1919.

We start by documenting broader patterns of economic and cultural assimilation for Jewish immigrants from 1900 to 1920. Two million Jews settled in the United States during this period, coming from Europe both for economic opportunity and to escape persecution. Using a new Jewish Names Index, we construct a new longitudinal dataset of Jewish immigrants in the US from 1900 to 1920. We document that Jewish immigrants had high occupation-based earnings relative to the US-born even upon first arrival, primarily due to their concentration in urban, semi-skilled jobs. Jewish immigrants experienced rapid earnings growth and significant cultural assimilation, as measured by names selected for their children, with time spent in the US.

We then ask whether the process of moving out of immigrant enclaves contributed to upward economic mobility. We focus on the IRO program, which encouraged poor Jewish immigrants to relocate from enclave neighborhoods to destinations around the country that were often pre-selected by program officers. Because Jewish enclaves in New York City were

---

[1] In 1910, the average immigrant from Southern or Eastern Europe lived in a neighborhood that was made up of at least 50 percent first- or second-generation immigrants, compared to only 10 percent of neighbors for the typical US-born household head (Eriksson and Ward, 2018, Figure 5).

[2] Jacob Riis, a photographer who publicized the conditions in the crowded tenements in New York City in *How the Other Half Lives* (1890), lamented that many of the children in immigrant neighborhoods "…had never seen the Brooklyn Bridge that was scarcely five minutes' walk away... The street, with its ash-barrels and its dirt, the river that runs foul with mud, are their domain." Henry Cabot Lodge, one of the lead champions of border restriction likewise emphasized that immigrants lived "in congested masses in our great cities. They furnish… a large proportion of the population of the slums" (Lodge, 1896).

characterized by overcrowding and concentrated poverty, the program combined relocation to neighborhoods with a lower co-ethnic share and a higher socio-economic status. Participants received moving expenses and short-term lodging in exchange for leaving New York City.[3]

Our analysis is based on newly digitized records for IRO program participants that we recovered from the American Jewish Historical Society. We compare IRO program participants to other Jewish immigrants who lived in the same set of enclave neighborhoods in the 1910 census and who arrived in the US in the same period. First, we link IRO participants and comparison men forward to the 1920 census to study economic and cultural assimilation. Then, we follow the sons of both groups to the 1940 census to examine intergenerational transmission.[4]

We find that immigrants who left a New York City enclave experienced faster earnings growth than their neighbors. Because the census did not collect income data before 1940, we compute a proxy for individual-level income ("income scores") based on detailed information on occupation, age, country of birth and state of residence. IRO participants earned 4 percent more by our income proxy in 1920 than a comparison group who held the same occupation and started in the same income quintile at baseline. Furthermore, these advantages persisted to the second generation, with the sons of IRO participants earning 7 percent more than the sons of non-participants in 1940.[5]

By leaving enclave areas, IRO participants also assimilated into the broader US society while retaining some of their Jewish identity. IRO participants married spouses with less distinctively Jewish first names, many of whom were probably from assimilated Jewish backgrounds but some of whom could have been non-Jews. Yet these couples did not select less Jewish names for their children, suggesting that men encountered a different pool of marriageable women outside of enclaves but nevertheless maintained their own cultural identity. Furthermore,

---

[3] Total monetary benefits of the program were small, the equivalent of around two weeks of pay for a low-skilled worker.

[4] Because women often change their surnames at marriage, we follow the literature by only attempting to link men who were moved through the IRO program and their sons.

[5] It is possible that the comparison households also benefited by having some men removed from their neighborhoods by the IRO program. However, relative to the size of Jewish enclaves in New York City at the time (750,000 Jewish residents), the program was small, representing only 5 percent of the population over a twenty-year period. Furthermore, if remaining residents did experience gains through lessened competition, we will under-estimate the benefits of leaving the enclave.

we find that men who themselves had distinctively Jewish names, implying a higher degree of association with Jewish culture, were most likely to move back to New York City after a period of relocation, suggesting that enclaves offered cultural amenities that some immigrants found to be a counterbalance to their economic costs.

Men who left New York City through the IRO program may have been different from their neighbors in unobservable ways – i.e., more resourceful or talented – and these attributes may have allowed them to move up the ladder even without mobility assistance. Although we lack a situation with random assignment into the IRO program (i.e., the program was never allocated by lottery), we provide suggestive evidence that participating in the IRO program conferred economic benefits. First, we compare IRO participants who were relocated earlier versus later in the program's history; men who moved earlier had more exposure to life outside of an enclave neighborhood by our follow-up date (1920). Men who moved earlier had the largest gain in income score, while men who moved latest in the program did not gain at all. We find no evidence that differential gains are driven by early movers being higher skilled at baseline, or having more family connections. Second, we compare IRO participants who moved with program assistance to men who left New York City enclaves on their own; these self-financed movers may have been selected for a similar set of attributes (e.g., risk tolerance) that encourage mobility. When we compare men with the same initial occupation, IRO participants and self-financed movers experience the same gains.

2. Contributions to Literature

Our findings suggest that leaving enclave neighborhoods facilitated immigrant advancement in the early twentieth century, but at the cost of lessening their cultural attachment. As a result, some immigrants who were more ethnically identified chose to move back to enclave neighborhoods. In addition, we offer a major advance on the historical literature on the Age of Mass Migration by producing primary data on a group – Jewish immigrants – who have heretofore not been separately identifiable in large datasets. Given the lack of information on religious affiliation in the census, studies of historical immigrant assimilation focus almost exclusively on country of origin, rather than religious group (Lieberson, 1980; Abramitzky, Boustan, Eriksson,

2014, Eriksson and Ward, 2018; Abramitzky, Boustan, Jácome, Perez, 2019; Ward, 2019).[6] Our new Jewish Names Index, which we will make broadly available to other scholars going forward, provides a window into historical Jewish communities.[7]

Our paper also contributes to the literature on the costs and benefits of living in immigrant enclaves. First, we provide some of the first evidence on the inter-generational effects of leaving immigrant enclaves. We follow the sons of participants into the labor market and find that the economic benefits of leaving an enclave persist to the second generation. Second, we focus on large enclaves that are more representative of immigrant neighborhoods today.[8] To date, the best causal evidence on immigrant enclaves is based on refugee assignment programs, mostly in Scandinavia (Edin, Fredriksson and Aslund 2003, 2004; Damm 2009, 2014; Beaman 2012). Refugee enclaves tend to be small: for example, the average refugee in Edin, Fredriksson and Aslund (2003) was settled in a municipality whose residents were only 8 percent foreign born overall (1 percent from the refugee's own home country). Refugees appear to gain from assignment to an area with others from their home country. This pattern has been interpreted as evidence that immigrant enclaves confer economic benefits more broadly.[9] By contrast, we document some economic costs of staying in large enclave areas.[10]

---

[6] The few studies that identify Jewish immigrants do so by using Yiddish speakers observed with the "mother tongue" question available in some census years (Pagnini and Morgan, 1990; Chiswick, 1983, 1991). Chiswick (1992) also uses historical information on Jewish immigrants in the Dillingham Commission report. Much of the literature on contemporary Jewish acculturation use the National Jewish Population Survey (Himmelfarb and Loar, 1984; Amyot and Sigelman, 1996; Kivisto and Nefzger, 1993; Burstein, 2007; Chiswick and Huang, 2008).

[7] Our new Jewish Names Index complements recent work by Zhang, Zuckerman and Obhukova (2016) and Fermaglich (2019) which analyze novel sources like World War I service records and name change petitions to document innovation and creativity in Jewish naming practices as a means to assimilate into US culture.

[8] Before relocation, the average IRO participant lived in a neighborhood that was at least 55 percent Jewish, on par with some large immigrant enclaves today (e.g., Mexicans in East Los Angeles).

[9] Beaman (2012) finds that recently arrived co-ethnics constitute labor market competition, whereas more established immigrants can provide assistance. Earlier work emphasizes the importance of immigrant enclaves in providing informal social insurance (Cohen, 1991), information about access to social services (Bertrand, Luttmer, and Mullianathan, 2000) and employment assistance (Munshi, 2003).

[10] Our results are consistent with recent historical work by Connor (2020) and Eriksson (Forthcoming) and with contemporary studies in the US (Borjas, 2000; Xie and Gough, 2011).

Finally, our paper contributes to the literature on mobility programs. We find a large out-migration response to a small financial incentive, similar to the effectiveness of small payments to encourage seasonal migration in Bangladesh (Bryan, Chowdury and Mobarak, 2014). However, many IRO participants eventually returned to New York City at a rate similar to the 10-15 year persistence rate in the Moving to Opportunity (MTO) study (Sanbonmatsu, et al., 2011, Exhibit ES-2).[11] Consistent with Bazzi, et al.'s (2019) study of a resettlement program in Indonesia, we find that leaving an enclave promotes cultural assimilation.

3. Patterns of Jewish Assimilation

More than two million Jewish immigrants moved from Europe to the United States during the Age of Mass Migration. The first large wave of Jewish migration from Germany in the 1860s was followed from 1880-1920 by poorer Jewish immigrants from the Russian Empire and other parts of Eastern Europe. Some Jewish immigrants were fleeing from anti-Jewish violence in Europe, while others were pulled to the US by economic opportunity (Kuznets, 1975; Boustan, 2007; Spitzer, 2013; Zipperstein, 2018). Jewish immigration slowed after the US border was restricted to new entry in the 1920s (Abramitzky and Boustan, 2017).

We start by documenting new facts about the economic and cultural assimilation of Jewish immigrants from 1900 to 1920. These facts rely on our new index of Jewish identity and on an "income score" variable that proxies for individual income. We explain the Jewish index and the income score in more detail in section 4.

First, Jewish immigrants had higher earnings ("income score") than the US-born even upon first arrival, primarily due to their concentration in semi-skilled urban occupations. Russian Jews experienced further earnings growth relative to the US-born with additional years spent in the US. We summarize these results in **Figure 1**, which presents coefficients from a regression of log income score on indicators for time spent in the US by country of origin or Jewish ethnicity. Coefficients are relative to US-born men, the omitted category. The panel sample of immigrants

---

[11] Immediately after receiving vouchers to move to low-poverty neighborhoods, MTO participants lived in census tracts with poverty rates 43 percentage points below the control group. However, after 4-7 years, the poverty rate gap between MTO compliers and control households fell to 20 points, and, after 10-15 years, the gap further declined to 10 points.

and US-born workers is observed in the 1900, 1910 and 1920 censuses (compare to Abramitzky, Boustan Eriksson, 2014; Figure 3; $N = 1.85$ million, with 44,000 likely Jews).[12]

Second, Russian Jews experienced the fastest cultural assimilation of any immigrant group during the Age of Mass Migration period. Here we define cultural assimilation as giving less foreign-sounding names to children born after spending more time in the US. **Figure 2** reports estimates by ethnicity or country of origin of the implied effect of spending 20 years in the US on the foreignness index of a child's name (compare to Abramitzky, Boustan and Eriksson, 2020; Figure 2 (Panel A); $N = 6.9$ million, with 406,000 likely Jews). Russian Jews are the group that shift their name choice the most with time spent in the US, with other Jewish immigrants in the middle of the pack.

Overall, Jewish immigrants arrived in the US with skills that allowed them to enter highly-paid occupations, and they continued to advance up the occupational ladder with time in the US. Our results are thus most relevant to high-skilled immigrant groups, of which there were many both in the past (see Figure 1) and the present (e.g., immigrants from China and India).

4. The Industrial Removal Office

Despite some economic successes, on average, many Jewish immigrants in the early twentieth century lived in enclave neighborhoods characterized by overcrowding and poor health conditions. Housing in the Lower East Side, in particular, was considered to be "stifling, unhealthy and miserable" (Maffi, 1995, p. 119). Together, charitable organizations like the B'nai Brith, the United Hebrew Charities and the Baron de Hirsch Foundation founded the Industrial Removal Office with the intention of "dispersing the immigrants [to] alleviate some of [the] problems [of]… filth, poor sanitation, disease, and soaring rates of delinquency and crime" (Rockaway, 1998, p. 1-

---

[12] The graph differs from Abramitzky, Boustan and Eriksson, 2014 (Figure 3) in four ways: first, we use the newly available complete-count census data, rather than a smaller sample initially compiled by the authors. Second, we use an improved crosswalk between original census occupation records and occupation-based income measures. Third, we define Jewish and non-Jewish immigrants in a mutually exclusive fashion, so that (for example) Russian, Austrian and German coefficients here are based only on non-Jewish immigrants. Fourth, our outcome variable is an "income score" estimated from the 1940 census, and applied to the men in our sample (1900-1920), rather than the 1950-based "occupation score" provided by IPUMS. See our website (https://ranabr.people.stanford.edu/matching-codes) for a replication of Abramitzky, Boustan and Eriksson (2014) using these improvements made since that paper was published.

3). Raising awareness for these efforts, Cyrus Sulzberger, the president of the Jewish Agricultural and Industrial Aid Society, addressed the National Conference for Jewish Charities in 1901 saying "go back to your communities and tell them…. to take these thousands of newcomers off New York's hands" (cited in Diner, 2000, p. 151).[13]

The first moves financed by the IRO program occurred in 1900. **Figure 3** graphs the number of IRO removals in each year of operation by country of birth. The program was most active from 1903 until the Panic of 1907, which led to a drop in overall immigration to the US and a decline in the willingness of communities around the country to accept and assist IRO participants. A second round of moves took place in 1912 and 1913. The program ceased operation after the closing of the US border to Eastern European migration in the 1920s.

The IRO targeted young Jewish immigrants experiencing economic hardship. Internal IRO documents reported that, in nine out of ten cases, applicants had experienced spotty employment for up to twelve weeks in the year before removal (IRO, 1911, p. 6). Participants learned about the program through public lectures, newspapers, referrals from other Jewish charities, or by meeting IRO agents that were stationed at Ellis Island. The IRO program maintained a storefront recruiting center in the Lower East Side. As an incentive for participation, the IRO offered moving expenses, as well as short-term lodging and help with job search at the destination. The average stipend for moving expenses was $15, the equivalent of around two week's pay for a low-skilled worker in the 1901 Cost of Living Survey.

Table 1 reports demographic and economic characteristics of the 39,000 household heads in the IRO records (we describe the dataset in more detail below). 79 percent of participants were men, most of whom moved alone, and the average age at removal was 28 years old. Around half of the cases were processed as "direct removals" comprising individuals with "no definite place to which they desire to be sent and who [left] the selection of the place to the judgement and discretion of the officials of the office." (IRO, 1911, p. 8). Other participants stated a locational preference – for example because they were moving to meet family.

---

[13] The IRO was one of many Jewish social assistance programs in New York; other agencies focused on poor relief and support for widows and orphans (Fridkis, 1981; Szajkowski, 1973). It also joined a wider effort aimed at *immigrant deflection* – as it was called – which sought to lessen overcrowding in New York City by re-routing Jewish immigrants through other ports of entry, including Galveston, TX (Eisenberg, 1995; Marinbach, 2012; Aaronson, Davis and Schulze, 2018).

We transcribed additional information from the IRO records for our complete linked sample of 2,362 individuals (the linking procedure is explained below).[14] At the time of their departure from New York City, 18 percent of participants reported having 'no trade,' a category that might reflect being an unskilled laborer. Other common occupations include semi-skilled positions like tailors, carpenters, blacksmiths and operators, which together represent around 25 percent of the sample. The majority of participants reported Russia as their country of birth (75 percent), with other Southern and Eastern European countries making up the balance. Relative to comparison households, IRO participants were somewhat more likely to be born in Russia and less likely to be born in Austria. We re-weight the data in our analysis to account for these differences in place of birth.[15]

IRO participants were assigned to more than 1,000 locations around the country, although participants were not compelled to stay in their assigned location, and our longer-term follow-up suggests that few of them did. Diner (2000, p. 152) summarizes these scattered locations, writing "The IRO sent Jewish immigrants to small communities – Champaign, Illinois; La Crosse, Wisconsin; Gary, Indiana; Galveston, Texas; Cedar Rapids, Iowa – all places quite unlike the Lower East Side in terms of Jewish numbers, density, and diversity. But the IRO also sent New York's Jewish newcomers to Cleveland, St. Louis and Chicago, places that had attracted immigrant Jews directly from eastern Europe."[16]

Figure 4 maps the share of IRO participants assigned to each location (Panel A). IRO participants were sent to the Midwest, the upper South and the Pacific states, with very few being resettled on the Eastern seaboard. Internal IRO documentation reports that around 90 percent of

---

[14] Data Appendix Table 1 explains how observations are lost in creating the linked sample. We drop women because they cannot be linked forward in the census. We also drop men with incomplete information on name and age, men whose names are below a certain threshold on the Jewish Names Index, men who are not unique in the 1910 census and thus cannot be matched forward, etc. Our final linked sample has 3,612 observations, of which around 65 percent have complete information on occupation and neighborhood.

[15] The place of birth distribution in our comparison sample is 65 percent Russian-born, 21 percent Austrian-born, 5.5 percent Romanian-born and 8.5 percent from other countries of origin.

[16] The IRO identified target locations through intermittent surveys and informal correspondence with established, but typically small, Jewish communities. Local contacts were charged with documenting the arrival of IRO participants, distributing stipends, and assisting with job search and accommodation. The extent of this assistance varied, depending on the in-kind offers of the local community.

participants were residing at the assignment location in the first year. However, by 1920, we find that only 10 percent of IRO households remain in the State Economic Area (county group) to which they were assigned.[17] The lack of "stickiness" in initial assigned location makes it impossible for us to use assigned location as an instrument for actual location, as in papers based on refugee assignments. Panel B shows that, by 1920, some portion of IRO participants had returned to the New York metropolitan area or other cities on the East Coast. However, in contrast to other Jewish households who had lived in an enclave in New York City in 1910 (Panel C), IRO participants remained more dispersed throughout the US, with large clusters in cities like Minneapolis, Denver, Portland and Los Angeles, that otherwise had yet to attract major Jewish settlement.

Figure 5 summarizes the effectiveness of the IRO program in removing participants from enclave neighborhoods. IRO participants were twice as likely as comparison households who lived in enclaves in New York City in 1910 to live outside of the New York area in 1920 (54 percent versus 28 percent), and 10 percentage points less likely to live in a Jewish enclave (defined here for descriptive purposes as an enumeration district that was at least 40 percent Jewish; results look similar using other thresholds). Note that enumeration districts have around 300 residents on average, around the size of a modern census block group. **Appendix Figure 1** graphs the full distribution of neighborhood Jewish share for IRO and non-IRO participants before and after relocation.[18]

Immigrant enclaves – both in New York City and other large metropolitan areas – were characterized by not only having a large foreign-born population but also having residents of lower socio-economic status. **Table 2** documents that – not surprisingly – immigrant enclaves in the 10 largest metropolitan areas had a higher immigrant share, but also fewer homeowners and fewer residents working in white collar positions. The IRO program was thus a "bundled" treatment,

---

[17] The IRO archive contains a number of disgruntled letters from participants complaining about their placement location. Rockaway (2018) quotes from this letter, dated August 23, 1905. *"Murderers! What did you want from us? Why did you send us to South Bend? We are going around hungry, and no work is found for us. We will die from hunger. The agent doesn't care at all. Now there is no other way for us, but to go straight to the river."*

[18] Before relocation, both groups were highly concentrated in neighborhoods that were above 60 percent Jewish. By 1920, many IRO participants had moved out of enclaves, whereas comparison households exhibited a bimodal distribution split between enclaves and integrated neighborhoods.

shifting participants to neighborhoods with fewer co-ethnics and more higher status neighbors. **Figure 6** confirms that, by 1920, IRO participants lived in neighborhoods with a lower Jewish share (by 7 percentage points) and also a higher white-collar share, a higher English speaking share and a higher homeownership rate.[19]

5. Data

We combine a series of historical sources to collect information on IRO participants before and after their relocation from New York City, and comparable information on non-participating households. We compiled the dataset in four steps: First, we identified IRO participants in the original program records, which were housed at and partially digitized by the American Jewish Historical Society. Second, we constructed comparison groups from the 1910 census of other Jews living in Jewish enclaves in New York City who did not participate in the program. Third, we linked IRO participants and comparison households forward to the 1920 census. We then link the children in these 1920 households forward to the 1940 census. Fourth, we define outcome variables, including measures of occupational mobility and cultural assimilation. We explain each step in turn.

5.1. The IRO records

We obtained records of IRO participants from the American Jewish Historical Society (AJHS), which digitized some of the information originally collected by the IRO program in order to facilitate genealogical research. In particular, the AJHS created an online searchable database with the following information on each program participant: first name, last name, year of removal, age at removal, and city of assignment. We augment this database by transcribing additional variables from the IRO ledgers for each participant, including birthplace, pre-participation occupation, and street address prior to leaving New York. **Figure 7** depicts an image of the records that we used to construct our dataset. Each of the variables that we added to the data is relevant for our analysis. We use birthplace as a characteristic in our census linking procedure (alongside name and age). Pre-program occupation allows us to examine who selected into the IRO program.

---

[19] Figure 6 is based on versions of equations (2) and (3), each using an enumeration district characteristic as our outcome variables.

Finally, we use street address to map participants to census geography in order to measure initial neighborhood characteristics and to find comparison households who lived nearby before removal.

We develop a geolocation procedure to map IRO participants into 1910 enumeration districts; a detailed description of this method is presented in Section 3 of the **Data Appendix**. Contemporary GIS software does not work well for this historical application given that street names, numbering systems and enumeration boundaries have changed across many US cities over the past century (Shertzer and Walsh, 2019; Connor, et al., 2019). Instead, we performed a fuzzy match between reported addresses in the IRO records and addresses in the 1910 census, which then allowed us to link each address to an enumeration district. Our method is similar in spirit to Akbar, Li, Shertzer and Walsh (2020) but was developed independently. In total, we match 71% of the street addresses in the IRO records. **Figure 8** identifies the 720 enumeration districts that define the boundaries of the four Jewish enclaves in New York City in 1910. On average, these districts were 44 percent Jewish by our measure, compared to the balance of enumeration districts in New York City, which were 6 percent Jewish.[20] This figure is a likely underestimate of Jewish population share given that our index will not capture all Jews (and very rarely classifies a non-Jewish person as Jewish; see our discussion below).

5.2. Constructing comparison groups using the new Jewish Names Index

Our main analysis compares IRO participants to other Jewish households who lived in a Jewish enclave in 1910. Because the census does not ask about religious affiliation, we identify likely Jews according to our new Jewish Names Index, which is based on the first and last name of household heads.

Our Jewish Names Index calculates the relative probability in the complete count censuses of 1920 and 1930 of a name (first or last) being held by a speaker of Jewish languages (Yiddish or Hebrew), relative to a speaker of non-Jewish languages. These relative probabilities are then normalized between zero and one for first and last names separately according to this formula:

---

[20] Diner (2000, p. 42) emphasizes that the boundaries of Jewish enclaves were not entirely clear. She cites the *WPA Guide to New York City* from 1939 as defining the neighborhood as "Fulton St. (South St. to Pearl St.) and Franklin St. (Baxter St. to Broadway) on the south to 14th St. on the north; from the East River west to Pearl St. and Broadway; excluding Chinatown."

$$Jewish\ Index_{name} = \frac{\frac{\#\ Yiddish\ speakers_{name}}{total\ \#\ Yiddish\ speakers}}{\frac{\#\ Yiddish\ speakers_{name}}{total\ \#\ Yiddish\ speakers} + \frac{\#\ nonYiddish\ speakers_{name}}{total\ \#\ nonYiddish\ speakers}}$$

Index values close to two (adding first and last name) are most associated with speakers of Jewish languages and names with values close to zero have no Jewish attachment. We then assign index values to all respondents in the 1910 census by first and last name (94 percent have a non-missing index value).[21] Our approach follows Fryer and Levitt's (2004) construction of a Black Names Index and Abramitzky, Boustan, and Eriksson's (2020) more general index of name foreignness.

We choose the threshold value of 1.4 on the Jewish Names Index above which individuals are considered likely Jews in order to match external estimates of the total size of the Jewish population from Kuznets (1975) (we consider robustness to alternative threshold values below). **Table 3** lists a set of names from the 1910 census that rank either very high or very low on the index, or around the threshold value. Individuals with traditional Jewish first and last names – such as Hyman or Abraham for first names and Cohen or Kaplan for last names – rank highly on our index. Individuals at the threshold have names like Jennie Snyder or Joseph Rich that could belong either to Jews or non-Jews.

We validate our index in the Canadian census of 1911, which includes information on both names and reported religion. Our index classifies as Jewish 53 percent of Jews and less than one percent of Christians (Catholics and Protestants). Thus, our index is very good at distinguishing Jews from non-Jews (low false positive rate), but is unable to capture all Jews in the population (some false negatives).

The Jewish Names index allows us to define a set of comparison households who lived in one of the four Jewish enclaves in New York City in 1910. Our comparison households must be likely Jews (index value > 1.4), and also male household heads, foreign-born, and between the ages of 16 and 49. Because our method will be less likely to identify Jews who were given (or selected for themselves) non-Jewish sounding names, we apply the same threshold value of the Jewish Names Index to our IRO sample as well for symmetry. **Data Appendix Table 1** reports

---

[21] Missing values occur because some individuals have only a first initial and because some very rare names are present in 1910 but not in the 1920 and 1930 censuses used to create the index.

the number of men in each subsample before the census linking procedure, including 25,130 IRO participants and 145,287 comparison households heads in a New York City enclave.

5.3. Record linkage

We estimate the effect of participation in the IRO program on later outcomes by following IRO participants and comparison households to subsequent censuses. We create two linked samples: one that links baseline measures to the complete-count 1920 census, and one that links sons observed in the 1920 households to the complete-count 1940 census. Baseline data is drawn from the IRO records in the year of removal (median year = 1907) for IRO participants and from the 1910 census for comparison households.[22] The datasets used in our matching procedure, and the match rates and sample sizes achieved are diagrammed in **Figure 9**.

Our matched samples are based on an automated algorithm developed by Abramitzky, Boustan, and Eriksson (2012, 2014) that creates links by first name, last name, age and state or country of birth.[23] For robustness, we also consider samples that apply the NYSIIS name standardization procedure before linking, or that take a more conservative approach and require individuals to be unique by name and birthplace within a five year age band.[24] Abramitzky, Boustan, Eriksson, Feigenbaum and Perez (2019a) demonstrates that these algorithms are on the frontier spanning the tradeoff between creating false links (type I errors) and missing true links (type II errors): the more conservative algorithm substantially reduces potential type I errors, and the less conservative algorithm reduces type II errors.

---

[22] We use the IRO records as a baseline observation for IRO participants rather than the 1910 census for two reasons. First, half of the removals took place after 1910, so many participants were not yet living in the US by the enumeration of the 1910 census. Second, finding IRO participants in the 1910 census would require that every IRO record is double matched (both to the 1910 and the 1920 census), which would limit sample size and would impose an asymmetric matching requirement on the treatment and comparison groups.

[23] The first step of the ABE algorithm screens the initial data for uniqueness by all linking attributes (first name, last name, age and country of birth). In our setting, we start by appending the IRO data to the 1910 census. We then create a sample that includes only unique observations, defined for IRO participants as being either (a) present as a singular observation in the IRO records or (b) present once in the IRO record and once in the 1910 census. Note that, because some IRO removals occur after 1910, we would not expect all IRO participants to be present in the 1910 census.

[24] Our main sample matches on exact name but first converts nicknames or name truncations to the likely extended version of the name (e.g. Abe to Abraham, Joe to Joseph).

Data Appendix Tables 1 and **2** report the sample sizes and match rates using alternative linking procedures. We link 3,612 (14 percent) of the IRO records to the 1920 US census and 27,904 (19 percent) of comparison households living in a Jewish enclave in 1910. These match rates are typical for foreign-born cases circa 1900.[25] We observe 4,285 sons living in IRO households in 1920, and 21,535 sons living in comparison households, and link 29 percent (31 percent) of these sons forward to the 1940 census.[26]

One concern with census linking is that it is easier to find a unique match for men who had an uncommon name or who reported an accurate age to the enumerator. Men with these characteristics often have higher socio-economic status than the general population (Abramitzky, et al., 2019a). **Data Appendix Tables 3** and **4** compare men in our matched sample to men in the IRO records (or to men in the 1910 census) who cannot be matched to 1920. Matched men score higher on the Jewish Names Index and our income proxy. To improve external validity, our main results are reweighted by baseline characteristics to match the full population. Column 4 in **Data Appendix Table 5** demonstrates that the reweighting procedure substantially balances the matched sample with the unmatched segment of the population.[27] We report unweighted results in the appendix.

5.4 Outcome variables for household heads and their sons

Table 4 includes summary statistics for IRO participants and other sets of comparison households from our linked sample. Recent immigrant arrivals were more likely to volunteer for relocation. The typical IRO participant in our sample arrived in the US in 1903, compared to an average arrival year of 1900 for other residents of Jewish enclaves and of 1896 for Jewish

---

[25] We suspect that the quality of the IRO records can explain the disparity in match rates between the IRO and the census samples, given that match rates for the sons linked from the 1920 census are more comparable (match disparity = 26 percent [=5/19] in the father generation and only 6 percent [=2/29] in the son generation).

[26] As is typical of historical linked samples, the match rates improve over census years, perhaps because the handwriting is easier for modern transcribers to decipher.

[27] Coefficients are weighted by the propensity of being matched $P_i(M_i = 1|X_i)$, which is calculated from a probit of match status on the covariates (e.g., age, farm status). Observations are reweighted by $(1 - P_i(M_i = 1|X_i))/P_i(M_i = 1|X_i) \times q/(1 - q)$, where q is the proportion of records linked.

households outside of enclave neighborhoods. We thus control flexibly for year of arrival in the US in our analysis. [28]

At the time of removal, IRO participants had substantially lower income scores than comparison households (earning $720 in 1940 dollars, relative to $990 for other enclave residents). Because the 1920 census does not contain individual earnings information, we use this income score as our main economic outcome. Our income score is based on a statistical model predicting income from covariates in the 1940 census (the first year with income data), and then using this model to assign income for men in earlier years. In particular, we regress log income in 1940 on fixed effects for 3-digit occupation, age, country of birth, and current state of residence, as well as all interactions.[29]

Both IRO participants and other residents of Jewish enclaves have similarly Jewishly-identified names by our Jewish Names Index (index = 1.83-1.84), whereas Jews that lived in other parts of New York City or in the rest of the country had less Jewishly-identified names (index = 1.76-1.77). All IRO participants and other enclave residents lived in the New York metropolitan area before removal.[30] By 1920, 13 years after the median removal, 54 percent of IRO participants remained outside of the New York area, compared to only 28 percent of comparison households. Before removal, IRO participants lived in similarly Jewish neighborhoods (54 percent versus 55 percent Jewish). By 1920, both participants and non-participants were living in more integrated neighborhoods, but this trend was particularly pronounced for IRO participants (a decline from 54 percent to 33 percent Jewish).

---

[28] Some of the difference in arrival years is mechanical, because all comparison households must have arrived by 1910 in order to be enumerated in the 1910 census, whereas some IRO participants arrived and were relocated after 1910.

[29] This method follows Abramitzky, et al. (2020) and is similar to the machine-learning approach for computing income scores proposed by Saavedra and Twinam (2018). Note that, for all interaction terms, we interact covariates with census division, instead of state. In addition, the 1940 census does not record farm income. We compute income for farmers following Collins and Wanamaker (2017). Specifically, we compute farmer incomes by multiplying the income of farm laborers in 1940 with the ratio of earnings for farmers versus farm laborers in the 1960 census, by region and immigration status. Few men in our sample are farmers.

[30] We use the State Economic Area boundaries to define living in the New York City area in 1920. This definition includes the five boroughs of New York City, along with Nassau, Rockland, Suffolk and Westchester counties.

6. Effects of participation in the IRO program

6.1 Estimation strategy

To study economic assimilation, we compare the income score of IRO participants to other residents of Jewish enclaves in New York City in the 1910 census, both before and after relocation. We stack data from two periods: before and after removal. Data before removal comes from the IRO records for program participants (median year = 1907) or from the 1910 census for comparison households. Post-removal observations are from the 1920 census. We then estimate:

$$y_{it} = \alpha + \beta_1(IRO_i x\ After_t) + \beta_2(IRO_i) + \beta_3(After_t) + X_i\Gamma + \varepsilon_{it} \qquad (1)$$

where the outcome variables *y* for household *i* include the logarithm of income score for fathers in 1920 or sons in 1940. The coefficient β_2 represents differences between program participants and comparison households before removal. We expect that $\beta_2 < 0$ if IRO attracted men who had poor labor market prospects. The coefficient β_3 represents income growth between 1910 and 1920; we expect our income score to be higher in 1920 ($\beta_3 > 0$). Our coefficient of interest is β_1, which tests whether IRO participants experienced greater earnings gains relative to non-participants after removal. If leaving New York led to improved earnings, we expect $\beta_1 > 0$.

Our preferred specification adds a set of fixed effects for initial occupation and for placement in the initial income score distribution (in quintiles), along with interactions with the $After_t$ indicator to allow for differential trends in earnings growth by initial economic characteristics.[31] These interactions are included in the vector X_i, along with fixed effects for a series of demographic attributes, including individual years of birth, individual years of arrival in the US, and birth place (Russian/not). In each case, we interact controls with the indicator for being in the 1920 census to allow for differential trends by group. In some specifications, we add a set of 1910 enumeration district fixed effects and their interaction with the $After_t$ indicator.

---

[31] We include 20 occupational fixed effects, one for each of the 19 most common occupations and then a 20th category for the remaining observations (which accounts for 16-18 percent of the data). Note that occupation and income score are not identical because the income score is also based on age, state of residence, and country of birth.

For our cultural assimilation measures, we observe the Jewish Names Index for a man's wife and children in 1920. We also compare the Jewish Names Index of participants and non-participants at baseline (circa 1910) to assess selection into the program on cultural attributes:

$$Own\ name\ index_{i1910} = \alpha + \beta_4 IRO_i + X_i \Gamma + \varepsilon_i \tag{2}$$

$$Wife/child\ name\ index_{i1920} = \alpha + \beta_5 IRO_i + X_i \Gamma + \varepsilon_i \tag{3}$$

β_4 measures whether IRO program participants had a stronger Jewish identity as measure by the Jewish Names Index, and β_5 indicates whether program participants were more likely to marry wives with distinctively Jewish first names or to give their children Jewish names. Equation (3) includes a control for a man's own Jewish Names Index in the vector *X*. Therefore, the coefficient β_5 reveals whether an IRO participant married a less Jewishly-identified spouse relative to a comparison individual with a similarly Jewish name. If living outside of New York exposed participants to a wider range of cultural influences and expanded their pool of marriageable women, we expect $\beta_5 < 0$.

6.2 Occupational attainment and cultural assimilation after removal

We now turn to understanding the effect of leaving enclave neighborhoods on economic and cultural assimilation. If living in an enclave hinders occupational attainment, we expect IRO participants to move up the occupational ladder after removal from the Jewish neighborhoods in New York City thereby increasing our income proxy. However, we note that participation in IRO entailed leaving a highly-paid state (New York) to move elsewhere in the country, which might dampen the effect of the program on our income score. Furthermore, if immigrants engage in social interactions primarily in their local neighborhoods, we would expect that IRO participants would encounter more assimilated (or even non-Jewish) spouses in their more integrated neighborhoods after removal. If retaining their cultural identity becomes less important to them, we would also expect IRO participants to choose less Jewish names for their children after removal.

We start in **Table 5** by comparing the income score of IRO participants and other residents of Jewish enclaves before and after removal. Consistent with the program's goals of assisting poor immigrants, individuals who availed themselves of the program had 18 percent lower earnings at

baseline (column 1). By 1920, around 10 years after removal, participants in the IRO program had converged almost completely with comparison households (column 2). As these two snapshots would suggest, the difference-in-differences coefficient implies that IRO participants gained 23 percentage points more in occupation-based earnings relative to comparison households in the decade after removal (column 3). This gain is apparent even after controlling for enumeration district (baseline neighborhood) trends, which is not surprising because we select our comparison households to reside in the same set of neighborhoods from which IRO participants are drawn (column 4).

Because IRO participants started out with lower earnings, the observed growth in earnings may simply reflect a process of convergence or economic assimilation driven by factors beyond neighborhood mobility. Our preferred specification in Column 5 thus compares men who held the same occupation and were situated in the same quintile of the income score distribution at baseline. A large portion of what appears to be gains from the IRO program was indeed driven by general convergence experienced by men in lower-paid occupations. Yet, we continue to find that IRO participants earned 4 percent more than others with the same socio-economic characteristics at baseline. We will use this specification for the rest of the paper.

Panel 2 of Table 5 follows families forward to 1940 – 30 years after the average removal – to observe their sons in the labor market at around age 30. In columns 1 and 2, we reproduce the father's regressions for men whose sons contribute to the analysis (that is, men who have sons in the 1920 census who can be followed forward to 1940). IRO participants were also negatively selected in this subsample – although slightly less so, earning 10 percent less than comparison households – and completely converged by 1920. In the 1940 census, we find that the sons of IRO participants in 1940 earned 7 percent more than the sons of comparison households whose fathers held the same occupation at baseline. Thus, the gains experienced by IRO households are retained into the second generation and possibly even expand, but the standard errors are large enough that we cannot rule out that the program gains are the same in the two generations.

Appendix Figure 2 more explicitly compares the intergenerational mobility of children of IRO participants with the children of non-participants whose parents had comparable baseline income as measured by their rank in the national income distribution (following Chetty et al. 2014 and Chetty et al. 2018, and Abramitzky et al. 2020). To do so, we first rank each son based on his income score relative to other sons born in the same year. Next, we rank fathers relative to all other

fathers with sons born in the same year. Ranks are based on position in the *national* income distribution. We then regress a son's rank on his father's rank, allowing both the slope and intercept to differ for sons of IRO participants and non-participants. We find a higher intercept for the sons of IRO participants relative to the sons of the non-participants, indicating higher levels of absolute mobility for IRO participants. Moreover, the slope for IRO participants is flatter than for non-participants, suggesting a weaker association between the initial rank of father and sons among IRO participants.[32] The intergenerational mobility associated with the IRO program is most apparent for families that started out below the median of the income distribution, suggesting that leaving enclaves allowed some families to move out of poverty.

IRO participants lived in more integrated neighborhoods in 1920, and thus may have had more interactions with non-Jewish neighbors, hastening the process of cultural assimilation. In **Table 6**, we ask whether IRO participants were less likely to marry Jewishly-identified spouses (as measured by Jewish Names Index), and to give their children Jewish sounding names, relative to comparison households who lived in Jewish enclaves in 1910. IRO participants were not selected on Jewish identity; rather, they held similarly Jewish-sounding names at baseline to comparison households. Yet IRO participants married spouses with less Jewish-sounding names, which may be an indication that exposure to life outside the enclave introduced them to a different pool of marriageable women. IRO participants and their spouses appear to select *more* Jewish-sounding names for their children, which is not consistent with the idea of changing cultural values. We will show below that these patterns are strikingly different by exposure to time spent out of New York.

6.3 Addressing selection into the program

Men who chose to participate in the IRO program may have been different from their neighbors in unobservable ways – i.e., more resourceful or talented – and perhaps these attributes, rather than the mobility assistance through the IRO program, can explain their occupational attainment after removal (akin to a classic Ashenfelter (1978) dip).

---

[32] **Appendix Figure 3** produces a similar graph for intra-generational mobility, wherein we plot a father's position in the income distribution in 1920 relative to his rank at baseline. We find that IRO participants were more likely to move up the distribution relative to non-participants, and that this mobility occurred most prominently below the median.

We address the potential for selection on unobservable attributes in a number of ways. First, we consider variation in exposure to life outside of New York City among IRO participants. If leaving enclave neighborhoods was salutary, we would expect that men who had longer exposure to life outside of the city by our follow-up year (1920) would experience the strongest economic benefits from initial removal. Exposure to life outside of New York can increase on two margins: removal from the city earlier versus later in the program, and/or remaining outside New York City rather than returning to the city. We consider each in turn.

In the first panel of **Table 7**, we split IRO participants into three groups of roughly equal size based on when they were removed from the city: early removals (1900-06), middle removals (1907-11) and late removals (1912-1919). The three groups look similar in their initial income scores, with each earning 15-22 percent less than comparison households, but gains from the program monotonically increase with exposure to time outside of New York. Compared to men who held the same baseline occupation and were placed in the same quintile of the income distribution, IRO participants who were removed early earned 18 percent more than comparison households by 1920, men who were removed in the middle earned 6 percent more, and men who were removed late did not gain at all, and in fact appear to fall behind in income score (column 3). Together, these coefficients imply a 0.5-1.1 percent gain in income score for each year spent out of the immigrant enclave.[33]

We interpret the fact that IRO participants who moved earlier experienced larger income gains as evidence that greater exposure to life outside of New York City yielded a positive return. However, men who joined the IRO program earlier may have been different in unobserved ways than men who joined later. For example, men who joined earlier may have had fewer family connections in New York City, or more family outside of the city. Alternatively, men who joined earlier may have moved out of a sense of adventure (not as a response to a negative shock).[34]

We find no evidence for these alternative explanations. We still find higher gains for early movers in **Appendix Table 1**, which controls for the number of likely Jews who shared an individual's surname (column 3) as a proxy for having family in New York, or directly adding

---

[33] The middle period in this analysis contains the recession years of 1907 and 1908, which may influence selection into the IRO program and later success. Results look similar when dropping these two years.

[34] Recall that we control for arrival year in the US, and so here we compare two men who arrived in New York in the same year but left via the IRO program in different years.

surname fixed effects (column 4). The same is true in **Appendix Table 2**, which restricts the IRO sample to participants who are identified in the records as "direct removals" (i.e., those who were *not* leaving New York to meet family but instead were placed in locations by program officers). Late movers do not appear to be especially disadvantaged; instead, we find that late removals had the highest income score at baseline (Table 7). Furthermore, the benefits of early moves appear even when we drop men who moved soon after arrival in the US who may join the program for different reasons – e.g., out of sense of adventure (**Appendix Table 3**). Without strong evidence for alternative explanations, we conclude that year of removal is likely driven by the idiosyncratic timing of negative shocks that might prompt men to leave the city.

Panel B instead compares the occupational gains of IRO participants who returned to New York versus those that remained out of the city in 1920. Initially, these men look identical on income score. Yet, here we find the counterintuitive result that men who remained outside of New York earned 3 percentage points more than comparison households, whereas men who returned to New York earned 5 percentage points more. However, we show below that this gain is driven by regional differences in earnings: New York state was a high-earning location in 1940 relative to the rest of the country. We introduce an alternative income score below that does not assign different income levels by state of residence and find that men who remained out of New York gained more in movement up the occupational ladder (see Section 6). All other results in the paper are robust to the use of this alternative score.

Table 8 explores the effect of exposure to time outside of New York on cultural assimilation. Panel A does not find a strong relationship between time spent out of New York and spouse characteristics or names selected for children. These patterns are more consistent with a change in marriageable pool than with a shift in cultural attitudes, which would have taken some time to occur.

Panel B shows that the decision to return to New York was highly selective on Jewish identity, which is an interesting finding in its own right. IRO participants who chose to come back to New York had names that were themselves 1.8 points higher on the Jewish Names Index than other enclave residents, and gave their own children substantially more Jewish names after return (4.3 points).[35] By contrast, IRO participants who remained out of the city looked similar to

---

[35] To put this magnitude in perspective, consider that, in 1910, men in enclave neighborhoods were married to wives that scored 9 points higher on the index than men outside of enclaves (Table 3).

comparison households at baseline but married spouses with less Jewish-sounding names (3.6 points). Despite marrying women with less Jewish names, these couples selected similarly Jewish names for their kids, suggesting again that leaving enclaves shifted the pool of potential spouses but did not change cultural attitudes.[36]

If all men who chose to leave New York City (including through the IRO program) were more entrepreneurial or resourceful in some way than those who remained, a better comparison for IRO program participants might be other voluntary movers. **Table 9** subdivides non-participants who lived in a New York City enclave in 1910 into two groups: men who voluntarily left the enclave by 1920 and men who remained in the enclave by 1920 (the omitted category). We document two patterns. First, the IRO program allowed men who were otherwise struggling (i.e., who had low initial income scores) to leave the city, whereas self-financed movers held similar income scores at baseline to the comparison group. Second, IRO participants who remained outside of enclaves in 1920 gained the same amount as self-financed movers, each earning 4-5 percent more than comparison households who held the same occupation at baseline (column 3).Together, these patterns suggest that relocation through the IRO program generated a "return to mobility" similar to the benefits received by other voluntary movers. What made the IRO program distinctive was that it provided the funding and coordination necessary to prompt poorer residents to leave the enclave.

7. Robustness

We make a number of decisions with our data in order to produce main results. This section tests the robustness of our findings to each of these choices.

We start in **Appendix Table 4** by adding two comparison groups to our sample: Jewish residents of New York City who lived outside of enclave neighborhoods in 1910, and Jewish men outside of New York in 1910. At baseline, Jewish residents of non-enclave neighborhoods earned 10 percent more that comparison households and Jews outside of New York earned 13 percent less. Both groups converged with residents of enclaves (the omitted category) over time, but the

---

[36] Regressions underlying column 3 are estimated at the level of the individual child. As a result, men with multiple children in the 1920 household will enter the sample multiple times. Results look similar if we instead collapse the results to the level of the household head, instead using the average Jewish Names Index value as a dependent variable.

degree of convergence for IRO participants was faster. We report these results both weighted and unweighted by baseline characteristics.

Appendix Table 5 considers our definition of "likely Jews," which is based on a threshold value on the Jewish Names Index. Our main results use a threshold of 1.4. Here, we use two more conservative values, threshold values of 1.6 and 1.8, which respectively eliminate 10 percent and then 30 percent of the sample. We find a similar pattern of initial selection and earnings convergence for IRO participants in each case, with participants earning 4-5 percent more by 1920 than comparison men with the same initial occupation and income quintile.[37]

Our results are robust to our choice of matching algorithm. **Appendix Table 6** presents samples built using two alternative algorithms: a modified version of the ABE algorithm that standardizes names using the NYSIIS phonetic algorithm (rather than using exact names as recorded) and a more conservative version of the ABE algorithm that requires individuals to be unique by name and country of birth within a five-year age band. In both cases, we continue to find an earnings gain of 4-5 percentage points relative to men who held the same occupation and income quintile at baseline.

We next try dropping two groups of IRO participants in **Appendix Table 7**: men who have "no trade" listed in their occupational field (there is no equivalent "no trade" category in census data), and men who reported a preferred location to the IRO program officers. Neither subsample alters our results, with the remaining IRO participants earning 4-5 percent more than similar men from the comparison group. This finding suggests that the results are not being driven by the subset of IRO participants who left New York to join established family networks.

Appendix Table 8 considers two alternative income scores that do not vary by state of residence. Our main income score assigns different income levels to men who live in different states, according to the regional differences in wage levels in 1940. Here, we use the standard 1950 "occupation score" that assigns each individual the median earnings for his occupation from the 1950 census (panel A), as well as a modified version of our "income score" that allows earnings to vary only by occupation, age and country of birth (panel B). Because the variables have different scales, we compare across outcomes by considering the size of the percentage point gain (column 3), relative to the initial earnings gap (column 1). In both cases, the IRO program gain is twice as

---

[37] Results are also similar when we construct a Jewish Index based on last name alone because many Jews used common first names like Mary.

large as the main effect in Table 5, which incorporates the income losses associated with leaving New York. This pattern suggests that the true return to leaving enclaves alone (rather than combined effect of leaving an enclave neighborhood and a high-paying state) may be even larger than our main estimates suggest. **Appendix Table 9** uses our alternate income score to reproduce the results by year of removal (early/middle/late) and by return to New York (compare to Table 8). Most importantly, we find here that IRO participants who remained out of New York in 1920 enjoyed larger gains than participants who returned to the city (9 percentage points versus 4 percentage points), which supports the importance of exposure to time outside of the enclave.

8. Conclusion

Both today and in the past, many immigrants live in enclave neighborhoods, residentially segregated from the native born. We document the economic and cultural assimilation patterns of one such immigrant group during the Age of Mass Migration – Eastern European Jews – and study a unique program that relocated almost 40,000 Jewish households from enclave neighborhoods in New York City to more integrated areas around the country circa 1910. The Industrial Removal Office program provided the funding and coordination necessary to allow poorer residents to leave the enclave.

Overall, Jewish immigrants integrated into the broader economy and assimilated into society. We find that men who volunteered to be resettled through the IRO gained 4 percent more in income score by 1920 than comparison households that held the same occupation and income quintile at baseline, suggesting that leaving enclave neighborhoods contributed to this upward mobility. These benefits were transmitted to the next generation, as the sons of IRO participants earned more than the sons of comparable households in 1940. Ours is one of the first papers that documents the effect on adult outcomes for children that grew up in an immigrant enclave.

By leaving the large Jewish community in New York City, IRO participants were exposed to neighbors from more diverse backgrounds, and they married spouses with less distinctively Jewish names. However, these couples did not select less Jewish names for their children, emphasizing the importance that many immigrants place on retaining their cultural identity.

IRO participants who moved earlier in the program and thus spent more years out of New York City experienced the largest income score gains by 1920, and were especially likely to marry spouses with less distinctively Jewish names. Men who chose to eventually move back to New

24

York City were distinguished for having more Jewish names (a sign of cultural attachment). Returning to the enclave carried a cultural benefit through proximity with ethnic community, but may have come at an economic cost

Prior evidence from refugee resettlement finds that the small immigrant enclaves can be beneficial to their residents. By contrast, in the context of Jewish immigrants in the early 20th century, we document the economic costs of remaining in a Jewish immigrant enclave. This comparison raises the possibility that there is an "optimal" enclave size – namely, living with too few countrymen may limit ethnic networks, while living with too many may create isolation. Another possibility is that benefits of enclaves are heterogeneous across groups –unlike refugee populations, Jewish immigrants held relatively highly-paid occupations upon first arrival in the US. Understanding when and who are helped by enclave neighborhoods is a fruitful avenue for future research.

References

Aaronson, Daniel, Jonathan Davis, and Karl Schulze. 2018. "Internal Immigrant Mobility in the Early 20th Century: Experimental Evidence from Galveston Immigrants." Working Paper.

Abramitzky, Ran and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature* 55(4): 1311-45.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102(5): 1832-56.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122(3): 467–506.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez. 2019. "Automated Linking of Historical Data." NBER Working Paper No. 25825.

Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, and Santiago Pérez. 2019. "Intergenerational Mobility of Immigrants in the US over Two Centuries." NBER Working Paper No. 26408.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do Immigrants Assimilate More Slowly Today than in the Past?" *American Economic Review: Insights* 2(1): 125-141.

Akbar, Protoy A., Sijie Li, Allison Shertzer and Randall P. Walsh. 2019. "Racial Segregation in Housing Markets and the Erosion of Black Wealth." NBER Working Paper 25805.

Amyot, Robert P., and Lee Sigelman. 1996. "Jews without Judaism? Assimilation and Jewish identity in the United States." *Social Science Quarterly*, 177-189.

Åslund, Olof, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist. 2011. "Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy." *American Economic Journal: Applied Economics* 3(2): 67-95.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong. "Unity in diversity? how intergroup contact can foster nation building." *American Economic Review* 109.11 (2019): 3978-4025.

Beaman, Lori A. 2012. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from refugees resettled in the US." *The Review of Economic Studies,* 79(1): 128-161.

Bertrand, Marianne, Erzo FP Luttmer, and Sendhil Mullainathan. 2000. "Network effects and welfare cultures." *The Quarterly Journal of Economics,* 115(3): 1019-1055.

Borjas, George. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." American Economic Review, 85(3): 365-390.

Borjas, George. 1998. "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation." *Journal of Urban Economics*, 44(2): 228-253.

Borjas, George. 2000. "Ethnic Enclaves and Assimilation." *Swedish Economic Policy Review*, 89-122.

Boustan, Leah Platt. 2007. "Were Jews political refugees or economic migrants? Assessing the persecution theory of Jewish emigration, 1881–1914." *The New Comparative Economic History: Essays in Honor of Jeffrey G. Williamson*, 267-90.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh." *Econometrica,* 82(5): 1671-1748.

Burstein, Paul. 2007. "Jewish educational and economic success in the United States: A search for explanations." *Sociological Perspectives* 50(2): 209-228.

Catron, Peter. 2019. "The citizenship advantage: Immigrant socioeconomic attainment in the age of mass migration." *American Journal of Sociology*, 124(4), 999-1042.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–1623.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." Working Paper 24441, National Bureau of Economic Research.

Chiswick, Barry R. 1983. "The earnings and human capital of American Jews." *Journal of Human Resources*, 313-336.

Chiswick, Barry R. 1991. "Speaking, reading, and earnings among low-skilled immigrants." *Journal of labor economics,* 9(2): 149-170.

Chiswick, Barry R. 1992. "Jewish immigrant wages in America in 1909: An analysis of the Dillingham Commission data." *Explorations in Economic History,* 29(3): 274-289.

Chiswick, Barry R., and Jidong Huang. 2008. "The earnings of American Jewish men: Human capital, denomination, and religiosity." *Journal for the Scientific Study of Religion* 47(4): 694-709.

Cohen, Lizabeth. 1990. *Making a New Deal: Industrial Workers in Chicago, 1919-1939*. New York: Cambridge University Press.

Collins, William J. and Marianne H. Wanamaker. 2017. "Up from Slavery? African American Intergenerational Economic Mobility Since 1880." Working Paper 23395, National Bureau of Economic Research.

Connor, Dylan. 2020. "Class Background, Reception Context and Intergenerational Mobility: A Record Linkage and Surname Analysis of the Children of Irish Immigrants." *International Migration Review*, 54(1), 4-34.

Connor, Dylan. 2019. "The Cream of the Crop? Geography, Networks, and Irish Migrant Selection in the Age of Mass Migration." *Journal of Economic History*, *79*(1), 139-175.

Connor, Dylan, Myron Gutmann, Angela Cunningham, Kerri Clement and Stefan Leyk. 2019. "How Entrenched Is the Spatial Structure of Inequality in Cities? Evidence from the Integration of Census and Housing Data for Denver from 1940 to 2016." *Annals of the American Association of Geographers*, 1-18.

Damm, Anna Piil. 2009. "Determinants of Recent Immigrants' Location Choices: Quasi-Experimental Evidence." *Journal of Population Economics* 22(1): 145-174.

Damm, Anna Piil. 2014. "Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-Random Neighborhood Assignment of Immigrants." *Journal of Urban Economics,* 79: 139-166.

De la Roca, Jorge. 2017. "Selection in Initial and Return Migration: Evidence from Moves Across Spanish Cities." *Journal of Urban Economics* 100: 33-53.

Diner, Hasia R. 2000. *Lower East Side Memories: A Jewish Place in America*. Princeton University Press.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment." *The Quarterly Journal of Economics* 118(1): 329-357.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2004. "Settlement Policies and the Economic Success of Immigrants." *Journal of Population Economics,* 17(1): 133-155.

Eisenberg, Ellen. 1995. *Jewish agricultural colonies in New Jersey, 1882-1920*. Syracuse University Press.

Eriksson, Katherine, and Zachary A. Ward. 2018 "The Ethnic Segregation of Immigrants in the United States from 1850 to 1940." NBER Working Paper No. 24764.

Katherine Eriksson. Forthcoming. "Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants during the Age of Mass Migration" *European Review of Economic History.*

Fermaglich, Kirsten. 2019. *A Rosenberg by Any Other Name: A History of Jewish Name Changing in America.* New York University Press.

Fridkis, Ari Lloyd. 1981. "Desertion in the American Jewish immigrant family: The work of the National Desertion Bureau in cooperation with the Industrial Removal Office." *American Jewish History* 71(2): 285-299.

Fryer Jr, Roland G., and Steven D. Levitt. 2004. "The causes and consequences of distinctively black names." *The Quarterly Journal of Economics* 119(3): 767-805.

Himmelfarb, Harold S., and R. Michael Loar. 1984. "National trends in Jewish ethnicity: A test of the polarization hypothesis." *Journal for the Scientific Study of Religion*, 140-154.

Industrial Removal Office. 1911. *Tenth Annual Report of the Industrial Removal Office for the Year 1910*. New York.

Kivisto, Peter, and Ben Nefzger. 1993. "Symbolic ethnicity and American Jews: The relationship of ethnic identity to behavior and group affiliation." *The Social Science Journal,* 30(1): 1-12.

Kuznets, Simon. 1975. "Immigration of Russian Jews to the United States: Background and Structure". *Perspectives in American History,* 9.

Lieberson, Stanley. 1980. *A Piece of the Pie: Blacks and White Immigrants since 1880*. Univ of California Press.

Maffi, Mario. 1995. *Gateway to the Promised Land: Ethnic Cultures on New York's Lower East Side*. NYU Press.

Marinbach, Bernard. 2012. *Galveston: Ellis Island of the West*. SUNY Press.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics,* 118(2): 549-599.

Pagnini, Deanna L., and S. Philip Morgan. 1990. "Intermarriage and social distance among US immigrants at the turn of the century." *American journal of sociology* 96(2): 405-432.

Rockaway, Robert A. 2018. *Words of the uprooted: Jewish immigrants in early twentieth-century America*. Cornell University Press.

Saavedra, Martin and Tate Twinam. 2018. "A Machine Learning Approach to Improving Occupational Income Scores." Working Paper.

Sanbonmatsu, Lisa, Lawrence F. Katz, Jens Ludwig, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma K. Adam, Thomas McDade, and Stacy T. Lindau. 2011. "Moving to opportunity for fair housing demonstration program: Final impacts evaluation." US Dept of Housing and Urban Development.

Shertzer, Allison and Randall P. Walsh. 2019. "Racial Sorting and the Emergence of Segregation in American Cities." *Review of Economics and Statistics* 101(3): 415-427.

Spitzer, Yannay. 2015. "Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States, 1881-1914." Working Paper.

Szajkowski, Zosa. 1973. "The "Yahudi" and the Immigrant: A Reappraisal." *American Jewish Historical Quarterly* 63(1): 13-44.

Ward, Zachary. 2019. "The Low Return to English fluency during the Age of Mass Migration." *European Review of Economic History*.

Xie, Yu and Margaret Gough. 2011. "Ethnic Enclaves and the Earnings of Immigrants." *Demography*, 48, 1293-1315.

Zhang, Jiayin, Ezra W. Zuckerman, and Elena Obukhova. 2016. "A Lack of Security or of Cultural Capital? Acculturative Conservative in the Naming Choices of Early 20th Century US Jews." *Social Forces*, 94(4); 1509-1538.

Zipperstein, Steven. 2018. *Pogrom: Kishiev and the Tilt of History*. New York: W. W. Norton & Co.

Figure 1: Log income score of immigrants by country of origin, relative to the US-born, 1900-20
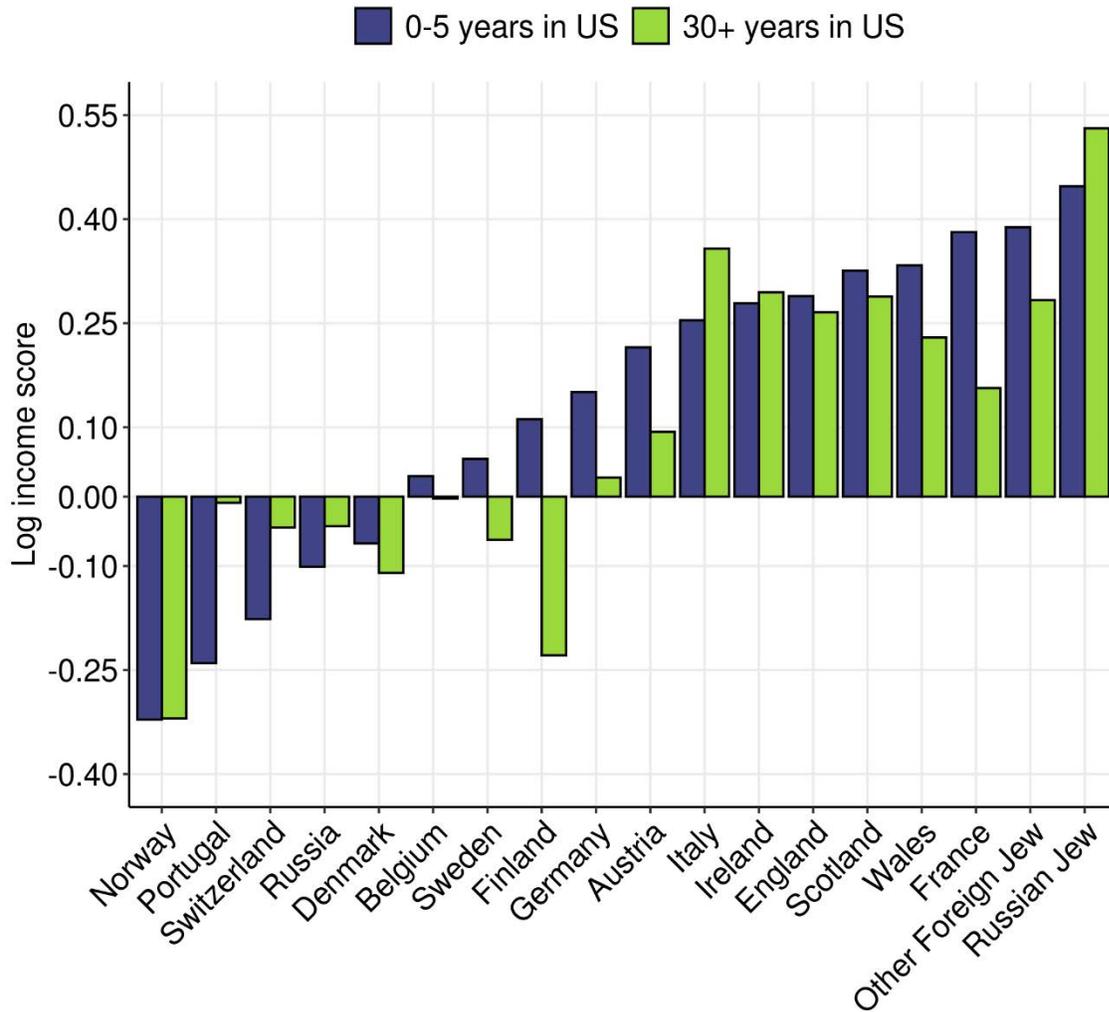
Notes: Log income score gap between the native- and foreign-born in a panel sample of immigrants and US-born workers observed in 1900, 1910 and 1920: The blue bars represent earnings gaps upon recent arrival (0-5 years in the US) and green bars represent earnings gaps after time in the US (30+ years in the US), by ethnicity or country of origin. This graph is a replication of Figure 3 from Abramitzky, Boustan and Eriksson, 2014 that distinguishes Jewish immigrants from other foreign-born. Jewish immigrants are separated into Russian Jews and Other Foreign Jews. At the same time, Jews are not included as part of other foreign-born countries of origin. The sample contains 1,854,029 observations out of which 43,708 have a Jewish name index > 1.4.

Figure 2: Changes in Foreignness Name Index of the second generation for each year immigrant mother spent in the US, 1920
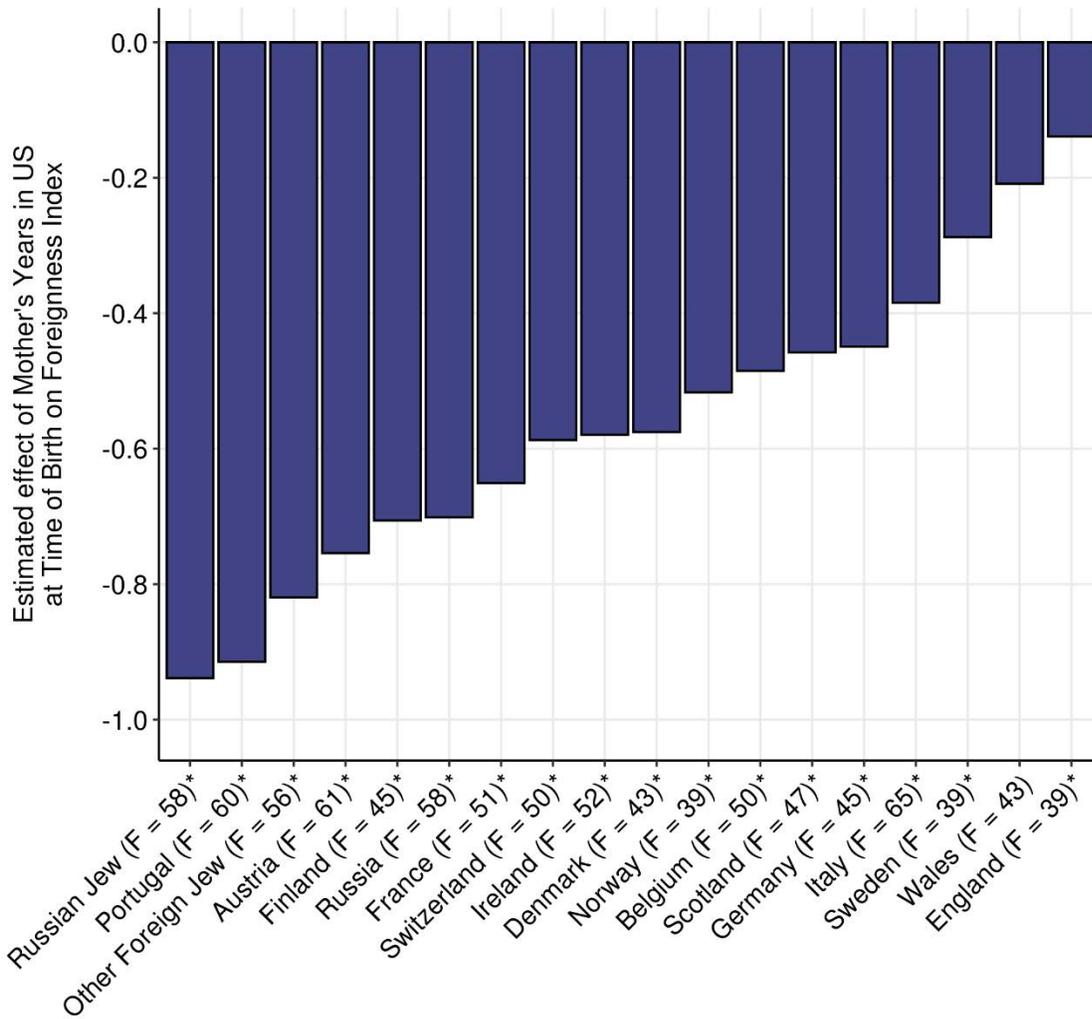
Notes: Effect of mother spending additional year in the US at time of child's birth on the foreignness index of a child's name, by ethnicity or country of origin. Estimates come from a regression of the foreignness index of a child's name on a set of interactions between mother's country of birth or ethnicity and years mother spent in the US at time of birth. This graph is a replication of Figure 2 (Panel A) from Abramitzky, Boustan and Eriksson, 2020 that distinguishes Jewish immigrants from other foreign-born. Jewish immigrants are separated into Russian Jews and Other Foreign Jews. At the same time, Jews are not included as part of other foreign-born countries of origin. The sample contains 6,945,895 observations out of which 406,369 have a Jewish name index > 1.4.

Figure 3: Frequency of Industrial Removal Office resettlements by birthplace, 1899-1919
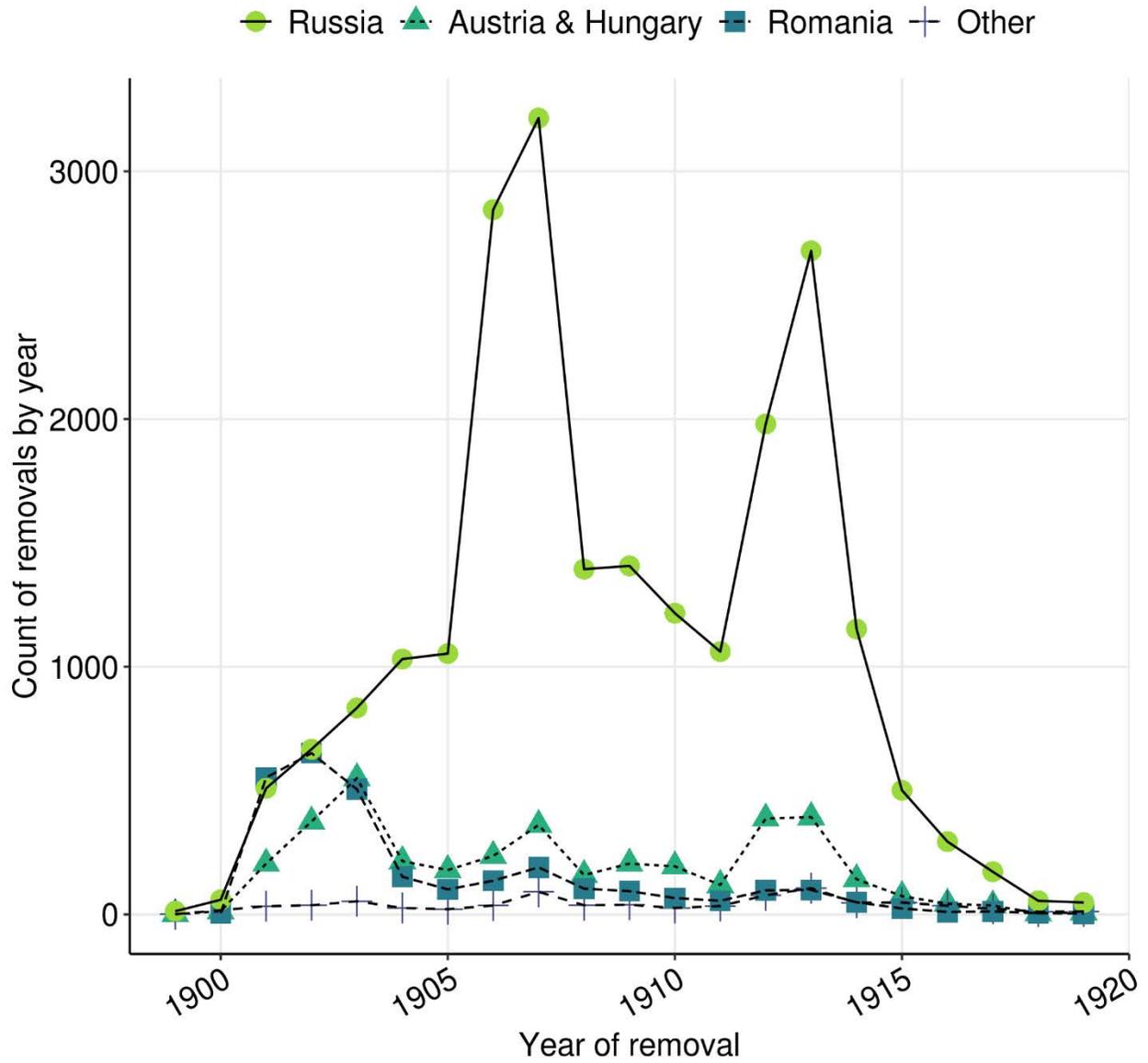
Notes: Yearly frequency of IRO resettlements between 1899 and 1920 by country of birth, based on the IRO record books.

Figure 4: Distribution of IRO and other foreign-born Jewish households across state economic areas, 1899-1920

A. Share of IRO assigned to location from 1899 to 1920
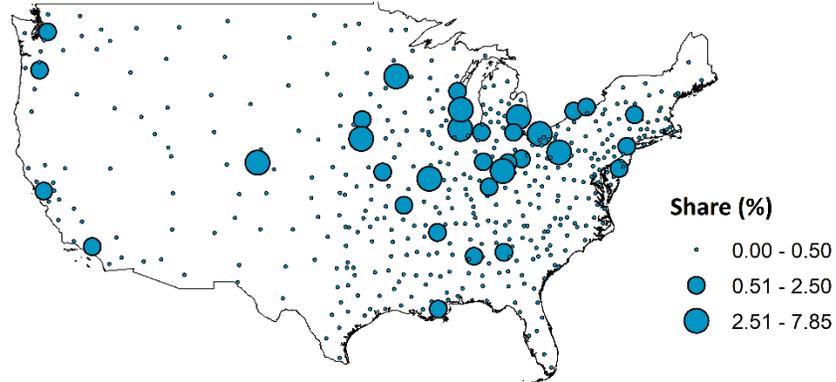
B. Share of IRO residing in location in 1920
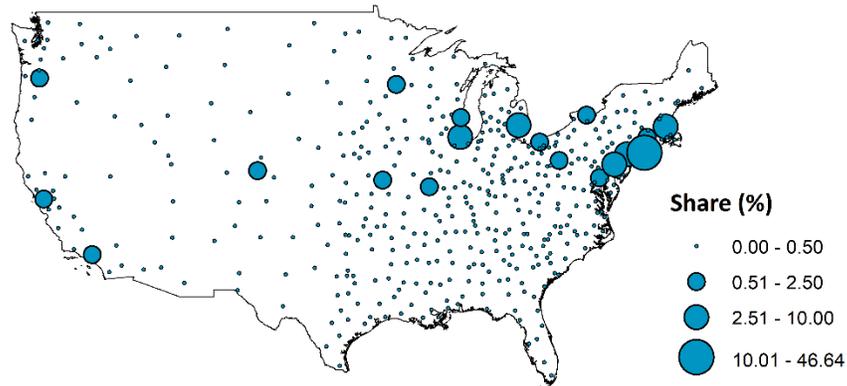
C. Share of other Jewish New Yorkers residing in location in 1920
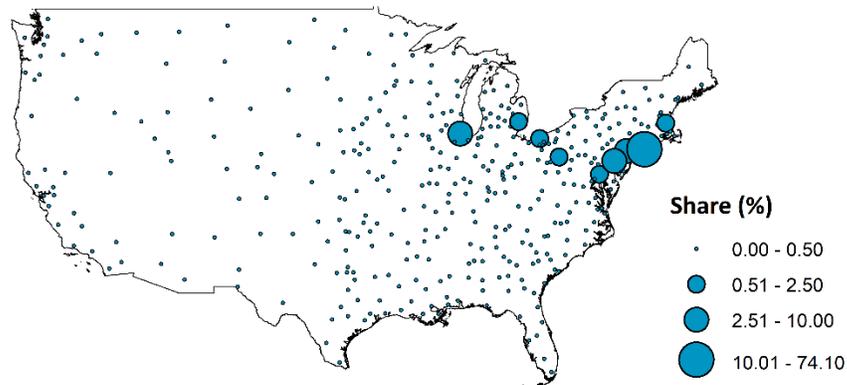
Notes: Panel A aggregates the placement cities reported by IRO to the scale of state economic areas (SEA) to display share (%) of IRO participants that were placed in different SEAs. Panel B uses the linked IRO-1920 Census sample to observe the 1920 post-resettlement locations of IRO participants. Panel C maps the locations of non-IRO Jewish New Yorkers, our main comparison group, from the 1920 census.

Figure 5: Outmigration rate based on IRO program participation, base period to 1920
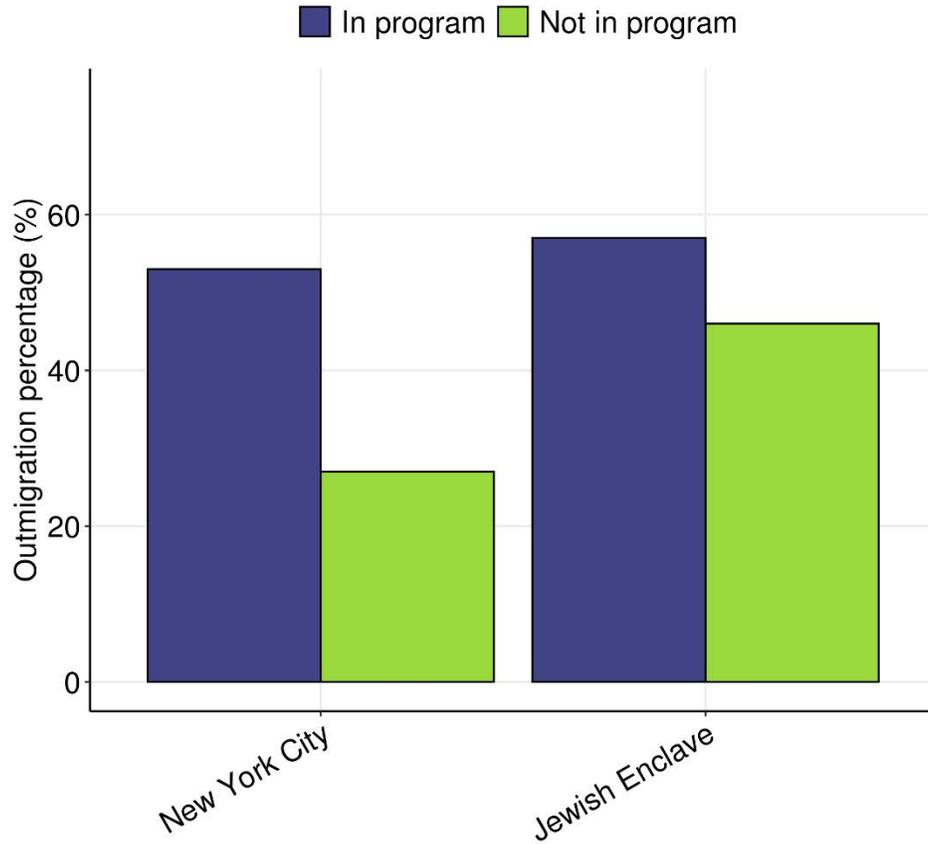
 Notes: Outmigration rate for IRO participants ("in program") relative to other Jewish immigrants ("not in program"). Both samples are restricted to men who lived in a Jewish enclave of New York City circa 1910 (see Figure 8). Outmigration is defined as living outside of the New York City state economic area in 1920 (left-hand bars) or living outside of a Jewish enclave in 1920 (right-hand bars). Enclaves are defined as enumeration districts anywhere in the United States that were at least 40 percent Jewish in 1920.

Figure 6: Comparison of IRO neighborhood characteristics, circa 1910 to 1920
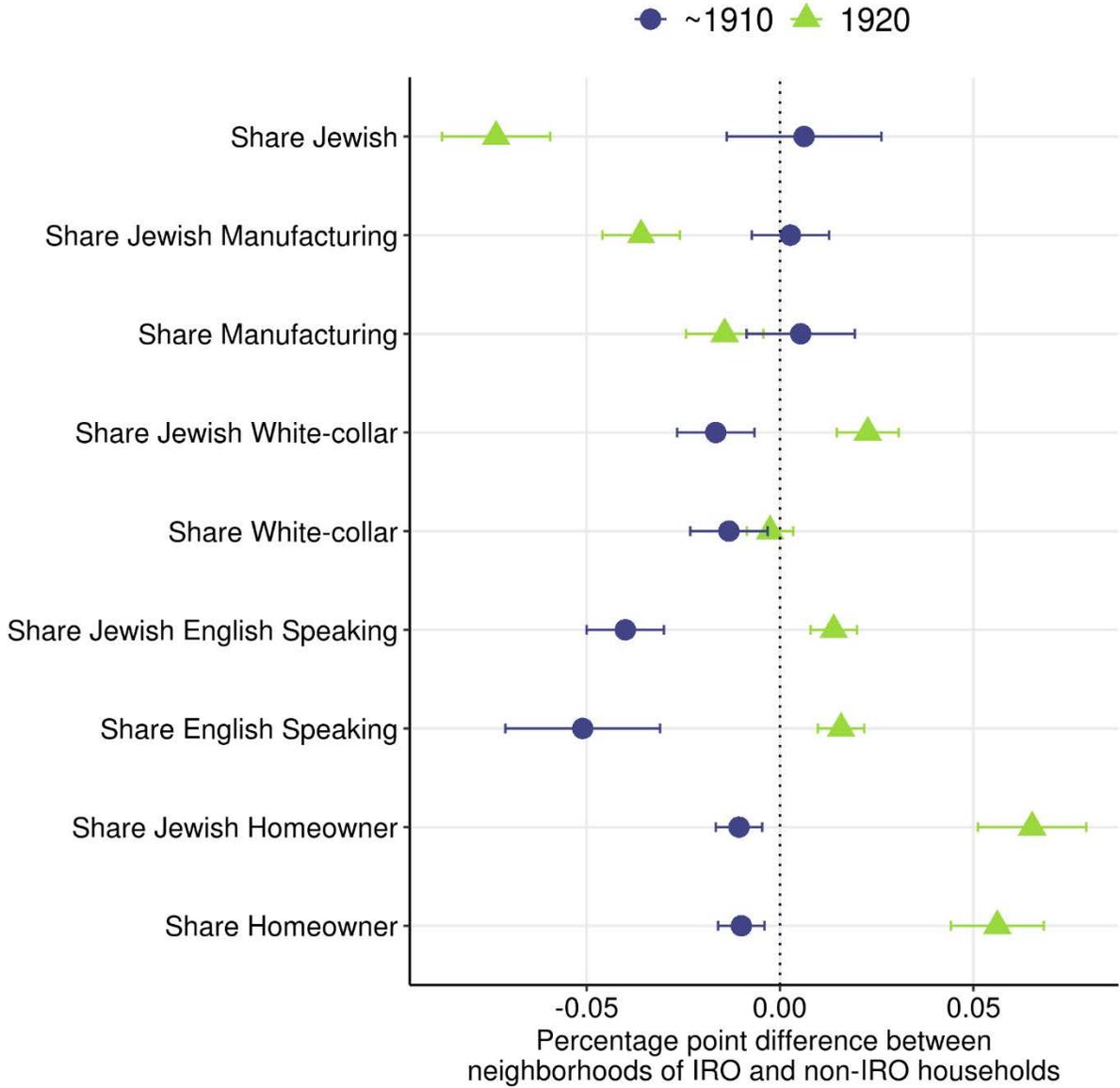
Notes: Percentage point difference between IRO and comparison group neighborhood attributes circa 1910 to 1920. The comparison group are Jewish households living in Jewish enclaves in New York in 1910. IRO sample is also restricted to those living in a New York enclave. The base period points represent coefficients and 95% CIs from nine separate regression equations using attributes from the 1910 Census as a dependent variable (version of equation 2). The 1920 points represent coefficients and 95% CIs from nine separate regression equations using neighborhood attributes from the 1920 Census as dependent variables and the 1910 enumeration district as a fixed effect (version of equation 3).

Figure 7: 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.

Figure 8: Delineation of New York Jewish enclave boundaries by the Jewish share of enumeration districts in 1910
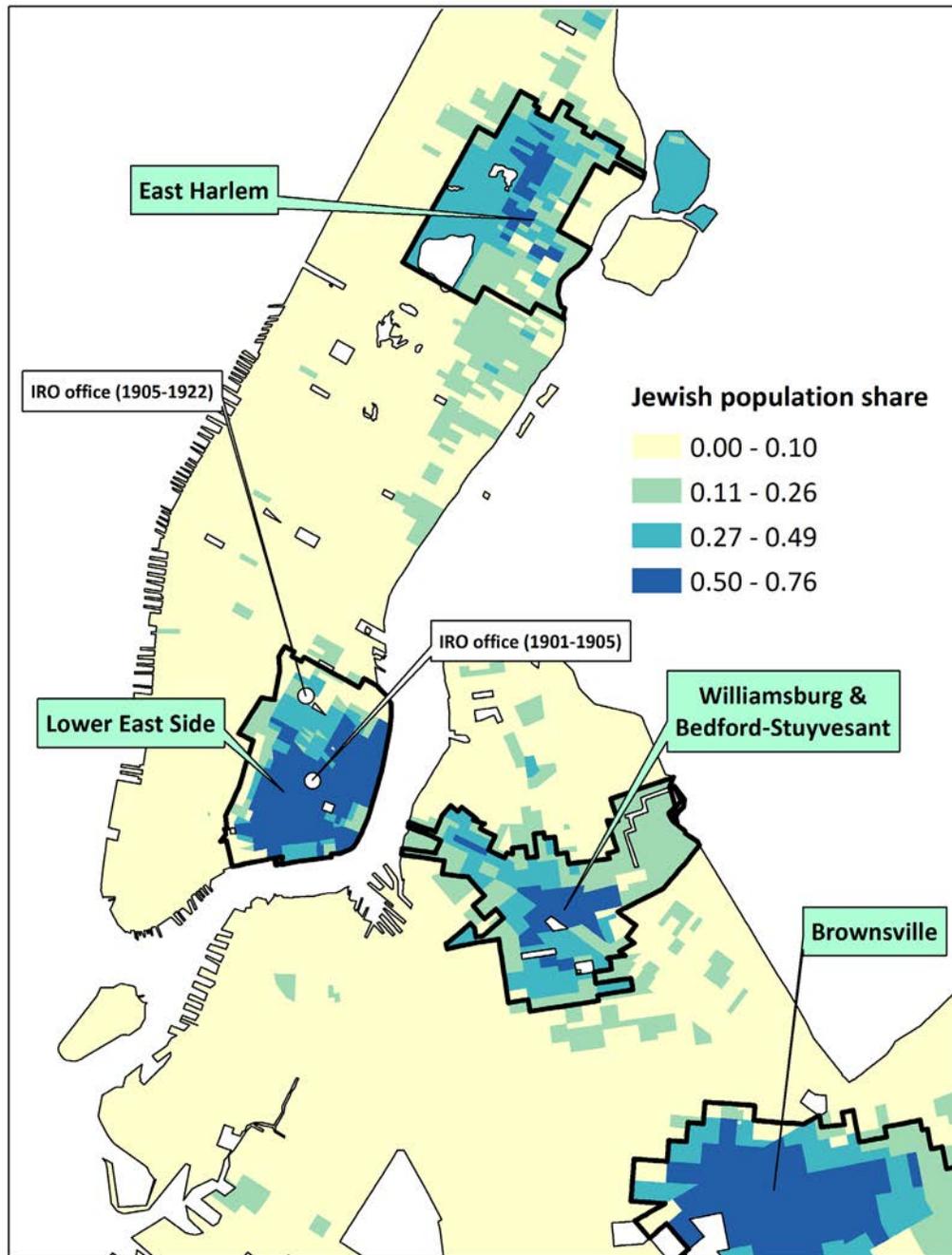
Notes: The boundaries of Jewish enclaves in New York superimposed on 1910 enumeration district boundaries. Boundaries of enclaves are determined by the Jewish population share of enumeration districts. The Jewish population share of enumeration districts is calculated from the share of individuals with a name-based Jewish index above 1.4. The black lines delineate the boundaries of Jewish enclaves. Allison Shertzer generously shared these digitized 1910 enumeration district boundaries.

Figure 9: Matching procedure and observation counts for IRO and preferred comparison (resident in New York enclave in 1910)

Industrial Removal Office Records	Base Period Sample	Preferred 1910 Census Comparison
First gen. men eligible for matching n = 25,130		First gen. men eligible for matching n = 145,287

Match rate = 14%   *Match rate = 19%*

1920 Census Household

First gen. matches n = 3,612	First gen. matches n = 27,904
Second gen. sons n = 4,285	Second gen. sons n = 30,768

Match rate = 29%   *Match rate = 31%*

1940 Census Household

Second gen. matches n = 1,241	Second gen. matches n = 9,385

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.

Tables

Table 1: Summary statistics for Industrial Removal Office participants

Dataset	Mean/share
A. Original IRO records (N= 39,004)	
Male	0.79
Travelled with wife	0.16
Direct removal	0.46
Mean age at removal	28 (8.79)
Mean arrival year in the USA	1903 (7.48)
B. Linked sample (N = 2,362)	
Top birthplaces	
Russia	0.74
Romania	0.10
Austria	0.08
Hungary	0.05
Turkey	0.01
Other stated birthplace	0.02
Top occupations	
No trade	0.16
Tailor	0.10
Carpenter	0.08
Operator	0.06
Painter	0.06
Other stated occupation	0.54

Notes: Descriptive characteristics of IRO participants from the transcribed IRO dataset. The original dataset included full transcriptions of name, age and year of removal. We transcribed birthplace – the only other essential characteristic for record linkage - for all participants. Following record linkage, we prioritized transcription of other attributes for linked cases (e.g. occupation, direct removal). Due to IRO data collection and our transcription procedure, we observe direct removal status for roughly 12,000 cases in total. Standard deviations are in parentheses.

Table 2: 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 (1940$)	772.78	820.57	804.32	880.07
Mean inc score, Jewish (1940$)	804.32	845.56	820.57	925.19
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.

Table 3: Jewish index for a sample of names held by over two hundred individuals in 1920

	Rank	First name	Last name	Observations (1920 census)	Jewish index
Most Jewish	1	Hyman	Levine	270	1.981
	2	Hyman	Goldberg	257	1.980
	3	Yetta	Cohen	378	1.980
	4	Hyman	Goldstein	229	1.979
	5	Hyman	Cohen	687	1.979
	6	Meyer	Cohen	335	1.976
	7	Isidore	Cohen	237	1.975
	8	Israel	Cohen	203	1.972
	9	Abraham	Shapiro	245	1.971
	10	Abraham	Kaplan	266	1.971
Borderline	957	Jennie	Snyder	363	1.402
	958	Anna	Winkler	235	1.401
	960	Sarah	Black	529	1.400
	961	Eva	Rose	211	1.400
	963	Gertrude	Schneider	214	1.398
	966	Joseph	Rich	274	1.395
Least Jewish	33,650	Francisco	Martinez	397	0.023
	33,651	Juanita	Martinez	222	0.022
	33,652	Maggie	Washington	230	0.022
	33,653	Guadalupe	Martinez	218	0.022
	33,654	Booker	Washington	248	0.021
	33,655	Clyde	Cox	222	0.019
	33,656	Floyd	Cox	230	0.019
	33,657	Clyde	Campbell	295	0.015
	33,658	Floyd	Campbell	235	0.015
	33,659	Wade	Hampton	219	0.011

Notes: Jewishness of a selection of the 33,661 names in the 1920 census held by at least 200 people. The counts by Jewish index are based on first and last name combinations. For example, there are 270 people named "Hyman Levine" and 363 people named "Jennie Snyder" in the 1920 Census.

Table 4: Summary statistics for IRO participants and various comparison groups

| | | Foreign-born, likely Jews | | |
	IRO	In NYC enclave, 1910	Out NYC enclave, 1910	Outside NYC, 1910
Demographic and economic				
Age (1920)	38	40	43	42
Arrival year	1903	1900	1896	1897
Income score, (~1910)	723.65	992.28	1234.29	992.11
Income score, (1920)	1270.45	1315.93	1427.55	1254.88
Second gen. income score, (1940)	1093.64	1192.61	1284.17	1109.04
New York resident (1920)	0.46	0.72	0.74	0.18
Lives in assigned SEA (1920)	0.10	-	-	-
Observations (N)	2,362	19,978	7,092	31,502
Cultural				
Jewish index of own name (~1910)	1.84	1.83	1.77	1.76
Jewish index of wife's name (1920)	0.73	0.73	0.63	0.63
Jewish index of child's name (1920)	0.63	0.57	0.50	0.50
Observations (N)	1,486	5,416	1,078	5,962
Enumeration district in IRO/1910				
Jewish share	0.55	0.55	0.14	0.16
Manufacturing share	0.46	0.45	0.25	0.26
White-collar share	0.27	0.30	0.45	0.34
English speaking share	0.50	0.54	0.74	0.68
Homeownership share	0.02	0.04	0.12	0.30
Observations (N)	1,491	19,976	7,088	31,471
Enumeration district in 1920				
Jewish share	0.33	0.39	0.25	0.19
Manufacturing share	0.30	0.33	0.27	0.25
White-collar share	0.36	0.35	0.40	0.36
English speaking share	0.75	0.75	0.80	0.79
Homeownership share	0.22	0.16	0.20	0.35
Observations (N)	1,491	19,976	7,088	31,471

Notes: Descriptive characteristics for primary samples from main analyses. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation circa 1910 and 1920. The smaller IRO samples for cultural and enumeration district attributes are a result of extra restrictions placed on these samples. For enumeration districts, the IRO sample is restricted to participants who lived in a New York enclave circa 1910. For cultural characteristics, individuals are restricted to men with no present spouse in the base period. The 1940 observations are based on second generation sons, for which 652 reported an income in the 1940 Census. Income scores are all denominated in 1940 dollars.

Table 5: Log income score of IRO participants in 1920 and second-generation sons in 1940

	Cross-section		Diff-in-diff		
	(1)	(2)	(3)	(4)	(5)
	~1910	1920	~1910-1920	~1910-1920	~1910-1920
A. First generation					
IRO	-0.180***	-0.0192**	0.226***	0.221***	0.0407***
	(0.007)	(0.008)	(0.011)	(0.012)	(0.009)
N	22108	22108	44216	43236	44216
	~1910	1940	~1910-1940	~1910-1940	~1910-1940
B. Second generation					
IRO	-0.109***	0.0307	0.140***	0.0371	0.0694*
	(0.012)	(0.037)	(0.039)	(0.062)	(0.041)
N	4554	4554	9108	8848	9108
Controls					
Birth cohort	Y	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y	Y
~1910 ED	N	N	N	Y	N
~1910 Occ.	N	N	N	N	Y
~1910 Inc. rank	N	N	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 (Columns 3-5) 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). Models including control for circa 1910 ED have fewer observations due to geolocation failure on IRO addresses.

Table 6: Cultural Assimilation of IRO participants, 1920

	Own Jewish index	Wife's Jewish index	Child's Jewish index
	(1)	(2)	(3)
	~1910	1920	1920
IRO	0.00776	-0.0180**	0.0185**
	(0.005)	(0.009)	(0.008)
N	6883	6883	12300
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
Own Jewish index	N	Y	Y
Child: age, sex, foreign born	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

Table 7: Log income score of IRO participants in 1920 by program exposure

	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
A. Years of treatment			
IRO: 14-20 years (early)	-0.184***	-0.0236	0.181***
	(0.027)	(0.027)	(0.015)
IRO: 8-13 years (middle)	-0.218***	-0.0135	0.0595***
	(0.011)	(0.012)	(0.012)
IRO: 1-7 years (late)	-0.147***	-0.0217	-0.0959***
	(0.017)	(0.017)	(0.013)
N	22108	22108	44216
	~1910	1920	~1910-1920
B. Compliance with relocation			
IRO: Returned to NYC	-0.175***	0.00205	0.0535***
	(0.009)	(0.010)	(0.011)
IRO: Stayed outside NYC	-0.184***	-0.0391***	0.0292**
	(0.009)	(0.011)	(0.011)
N	22108	22108	44216
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	Y
~1910 Inc. rank	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Table 8: Cultural assimilation of IRO participants in 1920 by program exposure

	Own Jewish index	Wife's Jewish index	Child's Jewish index
	(1)	(2)	(3)
	~1910	1920	1920
A. Years of treatment			
IRO: 14-20 years (early)	0.00986	-0.00568	0.00566
	(0.008)	(0.013)	(0.013)
IRO: 8-13 years (middle)	0.0136*	-0.0241*	0.0138
	(0.007)	(0.014)	(0.011)
IRO: 1-7 years (late)	0.000998	-0.0217*	0.0310***
	(0.008)	(0.012)	(0.011)
N	6883	6883	12300
	~1910	1920	1920
B. Compliance with relocation			
IRO: Returned to NYC	0.0179***	0.00196	0.0426***
	(0.006)	(0.011)	(0.010)
IRO: Stayed outside NYC	-0.00153	-0.0362***	-0.00109
	(0.007)	(0.011)	(0.010)
N	6883	6883	12300
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
Own Jewish index	N	Y	Y
Child: age, sex, foreign born	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 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.

Table 9: Comparing IRO participants to self-financed movers, log income score

	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
IRO, left enclave, 1920	-0.179***	0.00257	0.0526***
	(0.013)	(0.013)	(0.014)
IRO, in enclave, 1920	-0.174***	-0.0340***	0.00879
	(0.012)	(0.012)	(0.013)
Not IRO, left enclave, 1920	-0.00351	0.0344***	0.0401***
	(0.003)	(0.004)	(0.004)
N	22693	22693	45386
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	Y
~1910 Inc. rank	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Log income score by group and residence in enclave in 1920 (ED greater than 40 percent Jewish) for households that were resident in enclaves in 1910. Reference category are Jews that did not participate in IRO and lived in New York enclaves in 1910 and 1920. 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

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

Ran Abramitzky
Stanford University
& NBER

Leah Boustan
Princeton University
& NBER

Dylan Connor
Arizona State
University

June 2020

Appendix -- For On-line Publication Only

Appendix Figure 1: Jewish share of neighborhoods between IRO participants and Jewish households in New York enclaves circa 1910

Neighborhood Jewish share
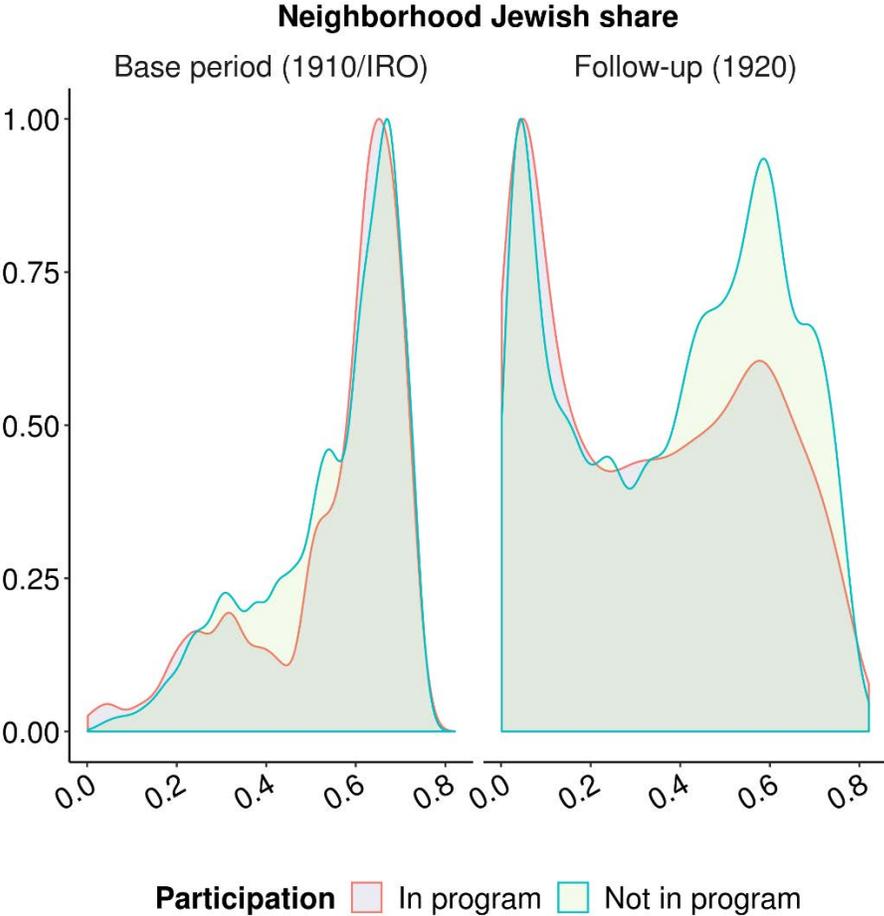
Notes: Kernel density plot of Jewish share of enumeration district circa 1910 and in 1920 for IRO and other Jewish households living in New York enclaves at baseline.

Appendix Figure 2: Rank-rank correlation for log income score of first-generation men in 1910 and in 1920
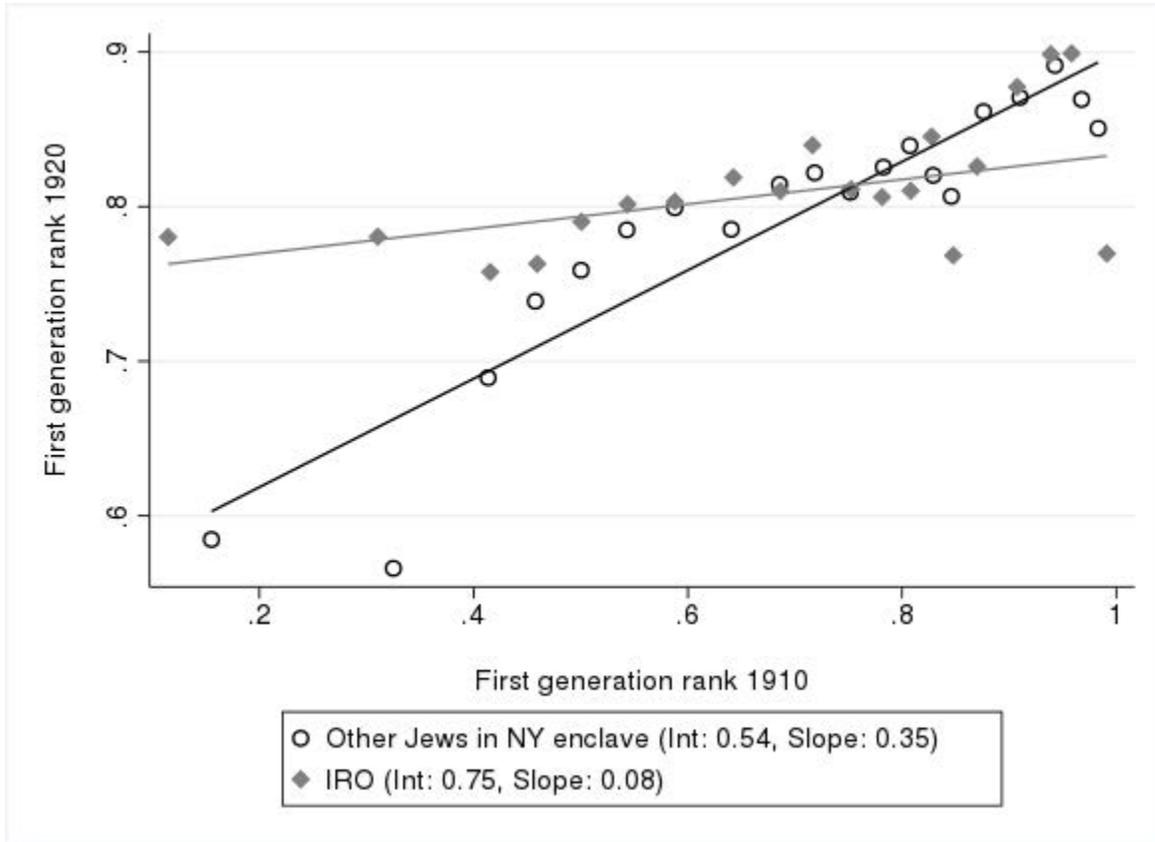
Notes: Binned scatterplot graphing the 1910- and 1920-income rank of IRO participants against other Jewish immigrants in New York enclaves (circa 1910). Men in each group are assigned percentile ranks based on their log income score. The figure plots the mean income rank for each group as well as the corresponding regression lines.

Appendix Figure 3: Rank-rank correlation for log income score of first-generation men in 1910 and second-generation sons in 1940
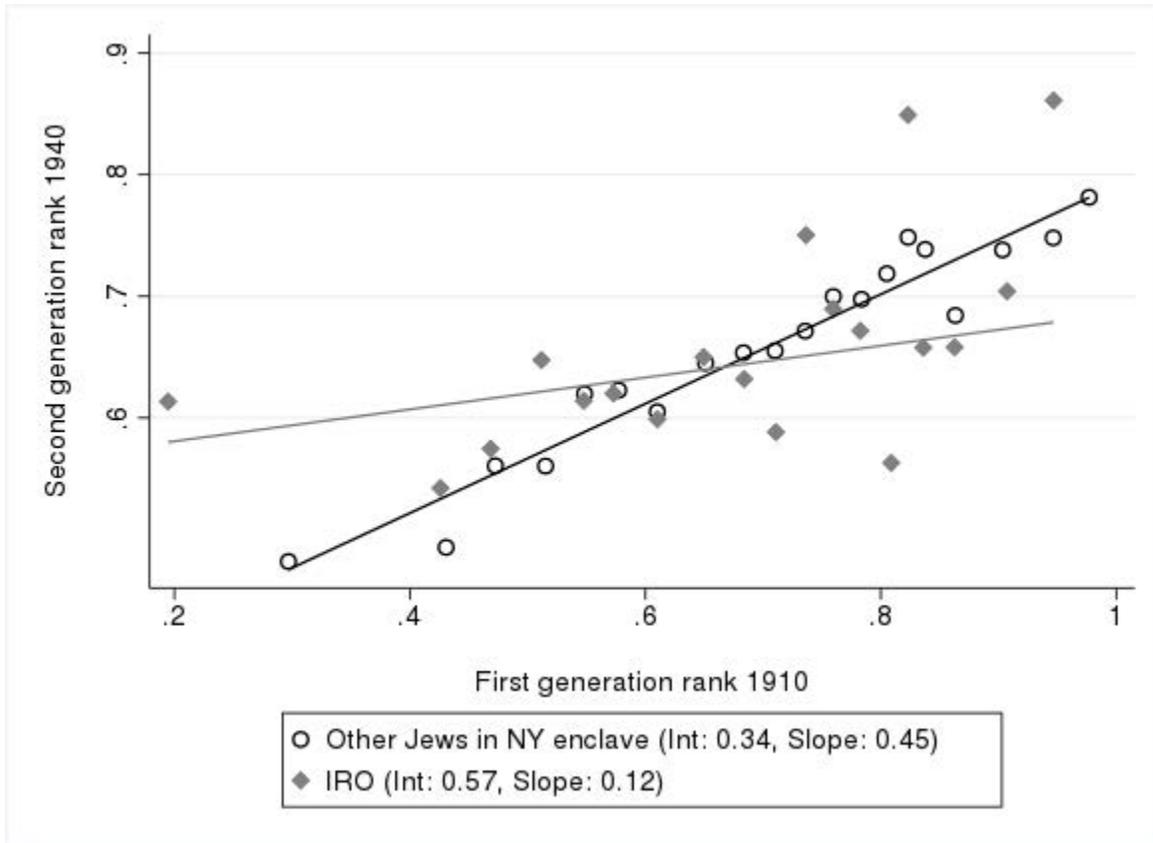
Notes: Binned scatterplot graphing the 1910 income rank of IRO participants and those of their sons in 1940 against the corresponding values for other Jewish immigrants in New York enclaves (circa 1910). The first and second generation in each group are assigned percentile ranks based on their log income score. The figure plots the mean income rank for each group as well as the corresponding regression lines.

Appendix Table 1. Log income score of IRO participants in 1920 by program exposure, with additional name-based controls

	Cross-section		Diff-in-diff	
	(1)	(2)	(3)	(4)
	~1910	1920	~1910-1920	~1910-1920
A. Years of treatment				
IRO: 14-20 years (early)	-0.180***	-0.0227	0.150***	0.121***
	(0.028)	(0.027)	(0.021)	(0.016)
IRO: 8-13 years (middle)	-0.217***	-0.0102	0.0654***	0.0688***
	(0.011)	(0.013)	(0.013)	(0.014)
IRO: 1-7 years (late)	-0.141***	-0.0139	-0.0737***	-0.0689***
	(0.017)	(0.017)	(0.013)	(0.015)
N	22035	22035	44070	44070
	~1910	1920	~1910-1920	~1910-1920
B. Compliance with relocation				
IRO: Returned to NYC	-0.173***	0.00822	0.0565***	0.0513***
	(0.010)	(0.010)	(0.013)	(0.012)
IRO: Stayed outside NYC	-0.178***	-0.0357***	0.0344**	0.0316**
	(0.009)	(0.011)	(0.014)	(0.013)
N	22035	22035	44070	44070
Controls				
Birth cohort	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y
~1910 Occ.	N	N	Y	Y
~1910 Inc. rank	N	N	Y	Y
Name-based network	Y	Y	Y	N
Last name	N	N	N	Y

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 (Column 3) and last name fixed effects (Column 4). 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 (Columns 3 and 4) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation).

Appendix Table 2: Log income score of IRO participants in 1920 by program exposure, with restriction to direct removals only

	Outcome: alternate log income score		
	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
A. Years of treatment			
IRO: 14-20 years (early)	-0.195***	-0.0420	0.138***
	(0.040)	(0.040)	(0.023)
IRO: 8-13 years (middle)	-0.232***	-0.0203	0.0513***
	(0.016)	(0.017)	(0.015)
IRO: 1-7 years (late)	-0.144***	-0.0442*	-0.117***
	(0.025)	(0.026)	(0.018)
N	20761	20761	41522
	~1910	1920	~1910-1920
B. Compliance with relocation			
IRO: Returned to NYC	-0.191***	-0.0221*	0.0188
	(0.013)	(0.012)	(0.014)
IRO: Stayed outside NYC	-0.187***	-0.0455***	0.00662
	(0.013)	(0.015)	(0.016)
N	20761	20761	41522
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	Y
~1910 Inc. rank	N	N	Y

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 restriction to direct removals only. Alternate log income score is based on the age adjusted average earnings of an occupation in the 1940 census and is not adjusted for birthplace or region of the United States. Reference category are Jews living in New York enclaves in 1910. The difference-in-difference coefficients (Columns 3 and 6) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation). Linear term for age at first observation included as additional continual control variable for IRO.

Appendix Table 3. Log income score of IRO participants in 1920 by program exposure, with IRO sample restricted to men who arrived in the US at least two years prior to removal

	Cross-section		Diff-in-diff
	Sample: IRO restricted to men living in the US >2 years before removal		
	(1)	(2)	(3)
	~1910	1920	~1910-1920
A. Years of treatment			
IRO: 14-20 years (early)	-0.178***	-0.0225	0.162***
	(0.034)	(0.034)	(0.019)
IRO: 8-13 years (middle)	-0.210***	-0.0326**	0.0316**
	(0.013)	(0.013)	(0.013)
IRO: 1-7 years (late)	-0.157***	-0.0367*	-0.100***
	(0.020)	(0.020)	(0.014)
N	21202	21202	42404
	~1910	1920	~1910-1920
B. Compliance with relocation			
IRO: Returned to NYC	-0.181***	-0.0103	0.0254**
	(0.011)	(0.010)	(0.012)
IRO: Stayed outside NYC	-0.178***	-0.0542***	-0.0104
	(0.010)	(0.012)	(0.013)
N	21202	21202	42404
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	Y
~1910 Inc. rank	N	N	Y

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 IRO sample restricted to men whose arrival year is more than two years prior to removal. 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 (Columns 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation).

Appendix Table 4: Log income score of IRO participants in 1920, comparison to Jewish men outside of enclave neighborhoods and outside of New York, weighted and unweighted estimates

	Estimates: weighted			*Estimates: unweighted*		
	Cross-section		D-in-D	Cross-section		D-in-D
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	~1910 -1920	~1910	1920	~1910 -1920
IRO	-0.189***	-0.0333***	0.0693***	-0.194***	-0.0464***	0.0689***
	(0.007)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)
Outside NYC Enclave, 1910	0.0988***	0.0628***	0.0206**	0.0969***	0.0547***	0.0104**
	(0.005)	(0.010)	(0.009)	(0.005)	(0.005)	(0.005)
Outside NYC, 1910	-0.133***	-0.0994***	0.0181	-0.134***	-0.0974***	0.0179***
	(0.005)	(0.012)	(0.012)	(0.003)	(0.003)	(0.004)
N	60214	60214	120428	60214	60214	120428
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
~1910 Occ.	N	N	Y	N	N	Y
~1910 Inc. rank	N	N	Y	N	N	Y

Standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01

Notes: Log income score differences between IRO and multiple comparison groups, weighted and unweighted. 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 (Columns 3 and 6) 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Appendix Table 5: Log income score of IRO participants in 1920, by Jewish name index threshold

| | Sample: Jewish index > 1.6 | | | Sample: Jewish index > 1.8 | | |
| | Cross-section | | Diff-in-diff | Cross-section | | Diff-in-diff |
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	~1910-20	~1910	1920	~1910-20
IRO	-0.188***	-0.0367***	0.0547***	-0.184***	-0.0246***	0.0389***
	(0.007)	(0.007)	(0.009)	(0.008)	(0.008)	(0.009)
Outside NYC Enclave, 1910	0.0920***	0.0482***	0.00923*	0.0922***	0.0441***	0.00716
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Outside NYC, 1910	-0.0948***	-0.0560***	0.0337***	-0.0557***	-0.0211***	0.0462***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
N	49061	49061	98122	34445	34445	68890
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
~1910 Occ.	N	N	Y	N	N	Y
~1910 Inc. rank	N	N	Y	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Log income score differences between IRO and multiple comparison groups, with more stringent thresholds for Jewish name index. 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 (Columns 3 and 6) 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Appendix Table 6: Log income score of first-generation men in 1920, matched using more conservative ABE algorithm and less conservative ABE NYSIIS algorithm

| | Sample: ABE NYSIIS | | | Sample: ABE CONSERVATIVE | | |
| | Cross-section | | Diff-in-diff | Cross-section | | Diff-in-diff |
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	~1910-20	~1910	1920	~1910-20
IRO	-0.192***	-0.0387***	0.0520***	-0.189***	-0.0653***	0.0433***
	(0.006)	(0.007)	(0.009)	(0.011)	(0.014)	(0.015)
Outside NYC Enclave, 1910	0.0972***	0.0490***	0.00746	0.0958***	0.0693***	0.0208**
	(0.004)	(0.007)	(0.007)	(0.006)	(0.009)	(0.009)
Outside NYC, 1910	-0.147***	-0.114***	0.00460	-0.133***	-0.0986***	0.0152
	(0.004)	(0.007)	(0.007)	(0.005)	(0.012)	(0.012)
N	75332	75332	150664	37875	37875	75750
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
~1910 Occ.	N	N	Y	N	N	Y
~1910 Inc. rank	N	N	Y	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Log income score differences between IRO and multiple comparison groups where observations are matched using NYSIIS names in the ABE matching algorithm and a more conservative ABE algorithm (Data Appendix). 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 (Columns 3 and 6) 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Appendix Table 7: Log income score of IRO participants in 1920, dropped observations where baseline occupations listed as "no trade" and restricted to direct removals only

	Sample: "No trade" dropped			Sample: "direct removals" only		
	Cross-section		D-i-D	Cross-section		D-i-D
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	~1910-20	~1910	1920	~1910-20
IRO	0.0516***	-0.00201	0.0524***	-0.200***	-0.0513***	0.0458***
	(0.001)	(0.002)	(0.010)	(0.010)	(0.013)	(0.015)
Outside NYC Enclave, 1910	-0.101***	-0.0269**	0.0184**	0.0987***	0.0630***	0.0217**
	(0.006)	(0.012)	(0.008)	(0.005)	(0.010)	(0.009)
Outside NYC, 1910	0.0979***	0.0550***	0.0187	-0.134***	-0.0993***	0.0205*
	(0.005)	(0.010)	(0.012)	(0.005)	(0.012)	(0.012)
N	52706	52706	105412	58867	58867	117734
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
~1910 Occ.	N	N	Y	N	N	Y
~1910 Inc. rank	N	N	Y	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Log income score differences between IRO and multiple comparison groups, with sample restricted to occupations other than "no trade" and to "direct removals". 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 (Columns 3 and 6) 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Appendix Table 8: Earnings of IRO participants in 1920, occupational scores (occ1950) and alternate log income score

	Outcome: Occupational score			Outcome: Alternate log income score		
	Cross-section		Diff-in-diff	Cross-section		Diff-in-diff
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	~1910-20	~1910	1920	~1910-20
IRO	-0.0950***	-0.000995	0.0472***	-0.266***	-0.00844	0.145***
	(0.006)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)
Outside NYC Enclave, 1910	0.0811***	0.0500***	0.0149	0.0927***	0.0590***	0.0218**
	(0.006)	(0.011)	(0.010)	(0.006)	(0.011)	(0.009)
Outside NYC, 1910	-0.0344***	-0.0176	0.0330***	-0.0668***	-0.0537***	-0.0132
	(0.004)	(0.011)	(0.012)	(0.006)	(0.012)	(0.013)
N	60214	60214	120428	60214	60214	120428
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
~1910 Occ.	N	N	Y	N	N	Y
~1910 Inc. rank	N	N	N	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Occupation-based earnings differences between IRO and multiple comparison groups by occupational score (OCC1950) and alternate log income score. Alternative log income score is based on the age adjusted average earnings of an occupation in the 1940 census and is not adjusted for birthplace or region of the United States. 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 (Columns 3 and 6) 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). Models including control for ED in base period have fewer observations due to geolocation failure on IRO addresses.

Appendix Table 9. Log income score of IRO participants in 1920 by program exposure, with alternate log income score

	Outcome: Alternate log income score		
	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
A. Years of treatment			
IRO: 14-20 years (early)	-0.290***	0.00845	0.211***
	(0.042)	(0.028)	(0.016)
IRO: 8-13 years (middle)	-0.314***	0.0110	0.0833***
	(0.017)	(0.013)	(0.012)
IRO: 1-7 years (late)	-0.197***	-0.00672	-0.0790***
	(0.025)	(0.017)	(0.013)
N	22108	22108	44216
	~1910	1920	~1910-1920
B. Compliance with relocation			
IRO: Returned to NYC	-0.250***	-0.0140	0.0383***
	(0.014)	(0.010)	(0.011)
IRO: Stayed outside NYC	-0.263***	0.0178*	0.0871***
	(0.013)	(0.011)	(0.012)
N	22108	22108	44216
Controls			
Birth cohort	Y	Y	Y
Arrival Year	Y	Y	Y
Russian birthplace	Y	Y	Y
~1910 Occ.	N	N	Y
~1910 Inc. rank	N	N	Y

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: IRO program exposure and alternate log income score changes by 1920 with IRO sample restricted to men whose arrival year is more than two years prior to removal. 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 (Columns 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation).

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"). 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	In NYC enclave, 1910	Out of NYC enclave, 1910	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	In NYC enclave, 1910	Out of NYC enclave, 1910	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%)
Matched with CONSERVATIVE ABE procedure (Match rate % to 1920 Census)	1,581 (7%)	17,026 (14%)	6,739 (18%)	27,849 (15%)

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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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

We classify Jewish households in New York City in 1910 by place of residence: inside or outside of 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). We were able to identify four Jewish enclaves depicted in Data Appendix Figure 1. These neighborhoods correspond to: Lower East Side; East Harlem; Bedford-Stuyvesant/Williamsburg; Brownsville. As the IRO primarily focused on moving Jewish households away from enclave neighborhoods, we select comparison households from the enumeration districts that made up these four enclaves in the 1910 census.
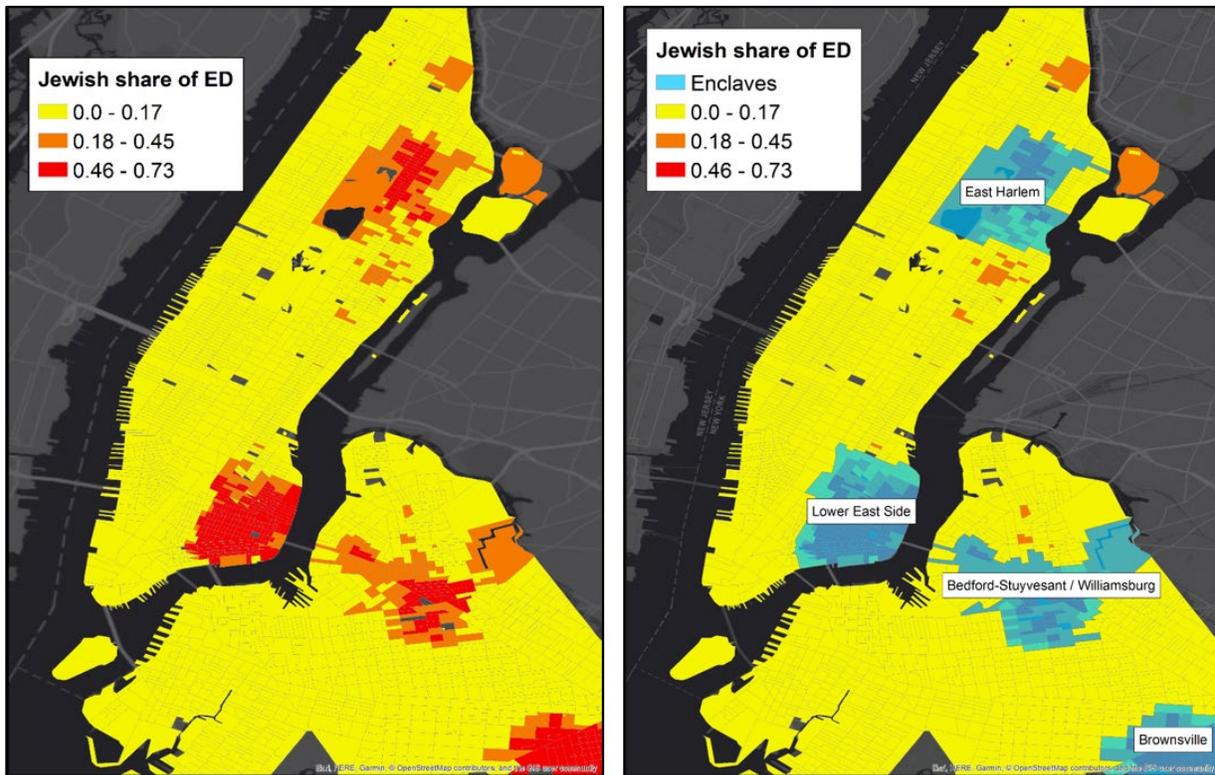
Data Appendix Figure 1. Jewish share of enumeration districts in New York in 1910 and neighborhood classification

Classifying the neighborhoods of IRO participants is more challenging. Although the IRO ledgers report a New York street address, these geographic identifiers are not easily geo-located or linked to the 1910 census enumeration districts. Given the historical changes in numbering systems, street names, and road networks of American streets, efforts to locate historical addresses using contemporary geocoders yield error rates of at least 30-40% (Connor et al., 2019). We suspect that the error rate in our case would be even higher because IRO addresses were recorded more than a century ago.

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, which we already coded as inside or outside of

Jewish enclaves. In particular, we use 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 relative 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