NBER WORKING PAPER SERIES

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

Ran Abramitzky
Leah Platt Boustan
Dylan Connor

Working Paper 27372
http://www.nber.org/papers/w27372

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020, Revised June 2023

We express our sincere gratitude to the American Jewish Historical Society and Susan Malbin for sharing the records of the Industrial Removal Office, and to David Rosenberg, Boni Koelliker and Tanya Elder for assistance with the archives. Keyoung Lee and Myera Rashid, along with Pritpal Araich, Katherine Delk, Jake Kantor, Conrad Makow and Alexander Newton, provided invaluable research assistance. Allison Shertzer generously shared her enumeration district shape files. Chris Becker, Alvaro Calderon, William Collins, Dora Costa, Martin Dribe, Katherine Eriksson, Nicky Halterman, Santiago Perez, Isaac Sorkin, Roger Waldinger and Maisy Wong offered helpful feedback, as did Jamie Goodwin-White, David Rigby and Michael Storper. We thank participants at meetings of the Association of American Geographers, the Homer Hoyt Institute, and the Population Association of America, and seminar participants at CUNY, LSE, Lund, the Minnesota Population Center, UCLA, Vanderbilt and Warwick. This research was supported by funding from the NUI Travelling Studentship, the Center for Jewish Studies at UCLA, the All-UC Economic History research group and a travel grant from the geography department at UCLA. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office

Ran Abramitzky, Leah Platt Boustan, and Dylan Connor

NBER Working Paper No. 27372
June 2020, Revised June 2023
JEL No. J15,N12,R23

ABSTRACT

The Industrial Removal Office funded 39,000 Jewish households to leave enclave neighborhoods in New York City from 1900 to 1922. Compared to neighbors with the same baseline occupation, program participants earned 4 percent more ten years after relocation. These gains persisted to the next generation. Benefits increased with more years spent outside of an enclave. Participants were more likely to speak English and married spouses with less Jewish names. More Jewishly-identified men (as measured by own name) were more likely to return to the city. We contextualize these results with new national evidence on Jewish economic and cultural assimilation.

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

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

Dylan Connor
School of Geographical Sciences
and Urban Planning
Arizona State University
Tempe, AZ 85282
dsconnor@asu.edu

Introduction

During the Age of Mass Migration (1850-1913), many immigrants to the United States lived in immigrant neighborhoods and relied on ethnic networks for social and economic support.[1] At the time, both pro- and anti-immigration voices expressed concerns about poor conditions in immigrant neighborhoods, arguing that the isolation of immigrant enclaves might impede assimilation (Lodge, 1909; Riis, 1890). 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 large-scale policy intervention by a non-government agency called the Industrial Removal Office (IRO) that financed 39,000 Jewish immigrant households to move out of Jewish enclaves in New York City between 1900-1922.

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, leaving Europe both to pursue economic opportunity and to escape persecution. Using both mother tongue (Yiddish) and a new Jewish Names Index to identify likely Jews in the US Census, we construct longitudinal data for 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 also

[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).

experienced 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 in New York City to destinations around the country. Jewish enclaves in New York City – like many immigrant neighborhoods at the time – were characterized by overcrowding and concentrated poverty. Thus, the program combined two features: relocation to neighborhoods with a lower co-ethnic share and relocation to neighborhoods with a higher socio-economic status. IRO also traded the access of participants to New York's opportunity-laden regional economy for a set of smaller, less dynamic labor markets that were scattered across the country (Connor and Storper, 2020). Participants received funding for moving expenses and train fare, as well as some short-term lodging and assistance in their new destination.[2] Although many participants did not stay in their initial assigned location and a sizeable minority even returned to New York City, we document that program participants were substantially less likely than others in their initial neighborhoods to live in a Jewish enclave ten years after resettlement.

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 held the same occupation at baseline. First, we link IRO participants and a group of comparison

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

households 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.[3]

We find that immigrants who left a New York City enclave through the IRO program experienced faster earnings growth than their neighbors who started with the same occupation. 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. We also allow these income scores to vary based on an interaction of region with occupation and country of birth, which helps incorporate New York's high intra-occupational earnings as well as the possibility that Jews were paid more in some labor markets than in others (Dillingham, 1911). IRO participants earned 4 percent more by our income proxy in 1920 than a comparison group. Furthermore, these advantages persisted to the second generation, with the sons of IRO participants earning 6 percent more than the sons of non-participants in 1940. IRO participants who settled outside of enclaves experienced the largest economic gains, on par with or more than neighboring residents who financed such moves on their own.

By leaving enclave areas, IRO participants also assimilated into broader US society while retaining some of their Jewish identity. IRO participants were more likely to speak English, which is found in some samples to contribute to immigrant economic advancement (Ward, 2020; Abramitzky et al., 2021a). They also 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

[3] 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.

that participants were able to retain their own cultural identity despite living in more integrated neighborhoods. Furthermore, we find that program participants who themselves had distinctively Jewish names were the most likely to move back to New York City. This pattern emphasizes that men with strong links to Jewish culture were particularly attracted to the ethnic and cultural amenities offered by Jewish neighborhoods.

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 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 by comparing sets of men within the program who ended up with different exposure to integrated areas. In particular, 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) and also experienced the largest gains among participants. We find no evidence that early movers had higher initial skills or more family connections.

We also analyze who chose to return to New York City after time spent away. Men who returned had more Jewish names at baseline, which may suggest stronger attraction to the cultural and religious amenities in the city. Men who returned were also more likely to give their children Jewish-sounding names. However, we find no selective return on initial earnings. Both men who remained out of New York City and men who returned experience economic gains, in part because the New York metropolitan area offered higher-than-average pay.

Our findings suggest that leaving enclave neighborhoods facilitated immigrant advancement in the early twentieth century with little cost in terms of lessening cultural attachment for the typical immigrant. The most ethnically identified immigrants did, however, chose to move back to enclave neighborhoods where cultural amenities were most plentiful.

Contributions to Literature

Our paper contributes to the broader literature on immigrant assimilation and the role of enclave neighborhoods in facilitating immigrant incorporation, as well as to the historical literature on the Age of Mass Migration.

First, we document that immigrants who leave a large enclave neighborhood experienced income gains and were more likely to engage in cultural assimilation. This finding contrasts with the existing literature on contemporary refugee assignment programs in Scandinavia, which finds economic gains associated with living near others from one's home country (Beaman, 2012; Damm, 2009; Edin et al., 2003).[4] Yet, refugees enclaves today tend to be far smaller than other immigrant enclaves. The neighborhoods that we study are more representative of large immigrant enclaves today.[5] Indeed, our results are consistent with recent historical work on the Irish (Connor,

[4] 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).

[5] The average refugee in Edin, Fredriksson and Aslund (2003) lived in a municipality where only 1 percent of residents were from the refugee's own home country. In our context, the average IRO

2020) and Norwegians (Eriksson, 2020), and with papers studying economic migrants in Germany, Australia and the United States today (Borjas, 2000; Danzer and Yaman, 2013; Laliberté, 2019; Xie and Gough, 2011).[6]

We provide some of the first evidence on the persistent inter-generational effects of leaving immigrant enclaves, following sons of IRO participants into the labor market. We also expand the analysis beyond economic effects to consider the cultural motivations for staying in enclaves and the cultural consequences for leaving immigrant areas. Ellis et al., (2004) and Bazzi et al. (2019) find that leaving enclave neighborhoods is associated with intermarriage and other markers of cultural assimilation.

Second, our work offers an advance on the historical literature on the Age of Mass Migration by producing primary data on a group – Jewish immigrants – who are hard to identify in large datasets (see Collins and Zimran, 2019; Connor, 2019; Spitzer, 2021; Xu, 2020).[7] Given

participant lived in a neighborhood that was at least 55 percent Jewish, on par with some of the largest immigrant enclaves today (e.g., Mexicans in East Los Angeles).

[6] O Gráda (2016) and Connor (2017) document that living in enclaves was associated with better health outcomes but lower literacy for Jews in Ireland.

[7] Most similar to our analysis is Spitzer (2018), who creates an algorithm to identify Jewish names using known Jewish arrivals in shipping records (classified as "Hebrew"). Collins and Zimran (2019) uses Irish Census data, which classifies respondents as Catholic and Protestant, to identify likely Catholic surnames. Xu (2019) separates ethnic groups among German, Russian and Polish immigrants in the US Census using a combination of name dictionaries, reported mother tongue, and common phonemes in ethnic languages.

the lack of information on religious affiliation in the census, many studies of historical immigrant assimilation focus on country of origin, rather than religious group (Abramitzky et al., 2014, 2021c; Eriksson and Ward, 2019; Lieberson, 1980; Ward, 2020).[8] In other work, Jewish immigrants have been identified indirectly using Russian birthplace or the reporting of Yiddish as mother tongue in the census (Chiswick, 1992, 1983; Pagnini and Morgan, 1990; Rosenthal, 1975).[9] The drawback to using birthplace to identify Jews is that many non-Jewish Russians will be misclassified as Jews, while the use of Yiddish mother tongue will undercount Jews that speak other languages. Our new Jewish Names Index provides a new approach to identifying Jews across censuses and generations, and will facilitate new research into historical Jewish communities. We report caveats for best use of this index below.

Third, our findings shed new light on the mechanisms supporting Jewish upward mobility in the historical United States. We provide further evidence on the role of ethnic enclaves in shaping immigrant attainment in the past. Analysis of data from the Dillingham Commission shows that foreign-born Jews earned 14 to 20 percent more than other immigrants in the early twentieth century, and they reached parity with native-born whites within four and half years of arriving in the United States (Chiswick, 1992). These outcomes partly reflect the relatively high

[8] Chiswick (1992) instead uses historical information on Jewish immigrants in the Dillingham Commission report.

[9] 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.

levels of skill that Jewish immigrants brought to the US (Kahan, 1978) and their disproportionate settlement in major immigrant gateway cities like New York (Chiswick, 1983). The benefits to living in New York were not just a feature of the broader regional labor market, but also ethnic networks that facilitated access to employment, self-employment, and training in lucrative areas of manufacturing work, such as garment making (Chiswick, 1992; Waldinger, 1986). Our findings suggest that the beneficial effects of living in New York may have been tempered by the large enclave neighborhoods there, and were likely higher for households who left these zones for more integrated parts of the metropolitan area (Abramitzky et al., 2021c; Connor and Storper, 2020).

Finally, our paper contributes to the broader 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 et al., 2014). The lack of stickiness of IRO participants in their original assigned locations is consistent with the Galveston Movement, a program that routed Jewish immigrants through the Port of Galveston and provided train tickets to preselected locations. Aaronson, Davis and Shulze (2020) find that more than 85 percent of Galveston participants left their original assigned location, often to move to large Eastern cities. The ultimate gains realized by the IRO participants is consistent with the view that the relatively high upward mobility rate of immigrants and children can partly be attributed to their weaker attachment to place, enabling flexibility in their search for opportunity (Abramitzky et al., 2021c).

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 (Abramitzky et al., 2021a; Boustan, 2007; Kuznets, 1975; Spitzer, 2021; 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 the next section later.

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 1A**, 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.[10] Coefficients are relative to US-born men, the omitted category. The panel sample of immigrants 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).[11]

[10] In Appendix Table 2, we document that the JNI performs very well in identifying Russian-born Jews by name.

[11] This figure updates the earlier graph by Abramitzky, Boustan and Eriksson, 2014 (Figure 3) by using: newly available complete-count census data; an improved crosswalk between original census occupation records and occupation-based income measures; defining Jewish and non-

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 1B** 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 et al., 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 being the third most rapid group.

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 analysis of the effect of enclave neighborhoods may thus be most relevant to high-skilled immigrant groups.

The Industrial Removal Office

Despite some economic successes, 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, 1994, p. 119).

Jewish immigrants in a mutually exclusive fashion, so that Russian, Austrian and German coefficients here are based only on non-Jewish immigrants; the "income score" estimated from the 1940 census rather than the 1950-based "occupation score" provided by IPUMS..

The Industrial Removal Office was founded by charitable organizations within the Jewish community – including the B'nai Brith, the United Hebrew Charities and the Baron de Hirsch Foundation – to alleviate these neighborhood conditions. The goal of the IRO program was to "dispers[e] the immigrants [to] alleviate some of [the] problems [of]… filth, poor sanitation, disease, and soaring rates of delinquency and crime" (Rockaway, 1998, p. 1-3). In "Dispersing the Ghetto", anthropologist Jack Glazier argues that the IRO emerged, in part, from cultural tensions between the German American and Eastern European Jewish communities. The architects of the program took the view that "Old World custom should adapt itself to the mores of the new country" and this adaptation process was being stymied by large immigrant enclaves (Glazier, 1998, p. 18-19). The IRO was thus framed as a means to accelerate the cultural assimilation of Jewish immigrants. 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" (Diner, 2000, p. 200, p. 151).[12]

The first moves financed by the IRO program occurred in 1900. **Figure 2** graphs the number of IRO program participants in each year of operation by their 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

[12] The IRO was one of many Jewish assistance programs in New York; other agencies focused on poor relief and support for widows and orphans (Fridkis, 1981; Szajkowski, 1973) and efforts which sought to lessen crowding in New York City by re-routing Jewish immigrants through ports like Galveston, TX (Eisenberg, 1995; Marinbach, 2012; Aaronson, Davis and Schulze, 2020).

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 (Industrial Removal Office, 1911, p. 6). Participants learned about the program through public lectures, newspapers or referrals from other Jewish charities. The IRO program also stationed agents to meet immigrants at Ellis Island and 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.

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 likely 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 linked the children in these 1920 households forward to the 1940 census. Fourth, we define outcome variables from the censuses, including measures of occupational and income score mobility and cultural assimilation. We explain each step in turn.

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. We present an image of the records that we used to construct our dataset in Appendix Figure 4. 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. 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 (Connor et al., 2019; Shertzer and Walsh, 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 et al. (2022) but was developed independently. In total, we match 71% of the street addresses in the IRO records.

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 develop an indirect approach to identify Jews in the data. Prior work shows that the majority of Jewish immigrants in the United States can be identified by whether they reported speaking "Yiddish" to the mother tongue question in the census (Chiswick, 1999; Rosenwaike, 1971). Unfortunately, the current version of the 1910 complete-count census – the base period for our comparison group – does not have a functional mother tongue variable.[13] To address this data limitation, we developed our own approach to identifying likely Jews in the 1910 census.

We classify likely Jews in the 1910 census by using the available information on Yiddish speakers in other proximate censuses (1920 and 1930) to identify first and last names associated with Yiddish speaking. In particular, 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:

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

[13] As of 2023, an error in the transcription of the 1910 complete-count census means that the mother tongue variable is unusable. This error does not affect the 1 percent sample of the 1910 Census.

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 foreign-born respondents in the 1910 census by first and last name (94 percent have a non-missing index value).[14] 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.

For our main analysis, we use a threshold value of 1.4 on the Jewish Names Index above which individuals are considered likely Jews. We selected this cut-off based on manual inspection in the overall population and around known Jewish neighborhoods. Importantly, the 1.4 threshold is the point at which 80 percent of Yiddish speakers in the 1920 census are classified as Jewish. We also show that our main results are robust to alternative threshold values below. **Table 1** 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 Harry Shaffer or Herman Schultz that could belong either to Jews or non-Jews.

With our index in hand, we can define Jewish enclave neighborhoods in New York City and Jewish household heads. We do so by using our names index to identify likely Jews in the 1910 Census. We then aggregate these person-level observations to calculate the Jewish population share of all New York enumeration districts, which we map in **Figure 3**. Note that enumeration

[14] 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.

districts have around 300 residents on average, around the size of a modern census block group. By identifying areas with clusters of disproportionately Jewish enumeration districts, we delineated the boundaries of the four Jewish enclaves in New York City in 1910 by hand: Lower East Side and East Harlem in Manhattan and Bedford-Stuyvesant/Williamsburg and Brownsville in Brooklyn.[15] On average, these districts were 44 percent Jewish by our names index, compared to the balance of enumeration districts in New York City, which were 6 percent Jewish. Our main analysis compares IRO participants to other male household heads who lived in one of these enclave neighborhoods in 1910 and who were also foreign-born, between the ages of 16 and 49, and likely Jews according to the names index.

One challenge in using this index to classify Jewish individuals is the fact that Jews make up a small share of the overall population. Classifying a small group leads to a well-known measurement problem that, even with a low *rate* of false positives (non-Jews classified as Jews), the overall sample can be overwhelmed by a high total *number* of non-Jews who exceed a given threshold (see, for example, Card, 1996 on classifying union membership). This problem mainly applies to the classification of Jewish individuals and is less relevant to the identification of

[15] 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."

distinctly Jewish neighborhoods, where aggregated counts of Jews dampens the noise introduced by false positives.[16]

We quantify this issue in **Appendix Table 2** by showing the agreements and disagreements in classifying likely Jews based on Yiddish mother tongue against our Jewish index in the 1920 Census (a year with accurate mother tongue data in the complete-count census files).[17] Yiddish was a mother tongue for many foreign-born Ashkenazi Jews. Thus, we can confidently assess the proportion of Yiddish speakers who are classified as non-Jews according to the JNI (false negatives). When we focus explicitly on individuals in our analysis sample: prime age immigrant men who lived in Jewish neighborhoods in New York at baseline, we correctly classify 84 percent of Yiddish speakers as Jewish by our index (false negative rate = 16 percent). For the same population, we also find that only 16 percent of men with a Jewish name by our index do not report speaking Yiddish. We suspect that some portion of this group tagged as "false positives" by this

[16] We validate this claim in **Appendix Figure 1** by comparing New York enumeration districts based on their share of Yiddish speakers in the 1920 census to their share Jewish, as implied by our names index. The correlation between these two measures is 0.85.

[17] As a second form of validation, we compare our Jewish Names Index values to known Jews in the Canadian census of 1911, which includes information on both names and religious affiliation. Our index classifies 53 percent of Jews and less than one percent of Christians (Catholics and Protestants) as Jewish. However, given the different sizes of the two populations (Jewish and Christian), our measure still implies an overall false positive rate over 40 percent.

metric were actually Jewish but spoke a different mother tongue (English, Russian, Polish, German) or misreported their mother tongue on the census.[18]

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 IRO records (median year = 1907) or 1910 census records to the complete-count 1920 census, and one that links sons observed in the 1920 households to the complete-count 1940 census.[19] The datasets used in our matching procedure, and the match rates and sample sizes achieved are diagrammed in **Appendix Figure 5**.

[18] In **Appendix Table 3**, we examine the sensitivity of our classification to different thresholds on the Jewish Names Index for the population of immigrants living in Jewish neighborhoods. The rate of possible false positives – non-Yiddish speakers classified as Jews by the names index – is very stable at around 25 percent across all thresholds of the JNI. The consistency of this rate provides further credence to our view that a large portion of this 25 percent of possible false positives are, in fact, Jews.

[19] 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.

Our matched samples are based on an automated algorithm developed by Abramitzky, Boustan, and Eriksson (2019, 2014, 2012), or "ABE", that creates links by first name, last name, age and state or country of birth.[20] Following Abramitzky et al. (2021b), we also consider samples linked using a variety of other criteria for robustness.

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.[21] 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.

[20] 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). To account for differences in name reporting across censuses, we standardize the shortened versions of names like "Abe" and "Joe" to "Abraham" and "Joseph". 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.

[21] 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).

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.[22] We report unweighted results in the robustness section below.

Effects of participation in the IRO program

Descriptive statistics for IRO participants and comparison households

Table 2 reports demographic and economic characteristics of the 39,000 household heads in the IRO records of which around 25,000 are eligible to be linked forward to the 1920 Census

[22] 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 $(1 - P_i(M_i = 1|X_i))/P_i (M_i =1|X_i)$ x $q/(1 - q)$, where q is the proportion of records linked.

(we describe the dataset in more detail below).[23] 79 percent of participants were men, most of whom moved alone, and the average age at removal was 28 years old. Nearly two-thirds 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." (Industrial Removal Office, 1911, p. 8). Other participants stated a locational preference – for example, because they were moving to meet family.

Our complete linked sample contains 3,612 observations (the linking procedure is explained below). We were able to transcribe additional information from the IRO records for 2,352 of these individuals. At the time of their departure from New York City, 16 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 30 percent of the sample. The majority of participants reported Russia as their country of birth (74 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.[24]

[23] **Data Appendix Table 1** explains how observations are lost in creating the linked sample, including the dropping of women, individuals with incomplete information on name and age, men whose names are below a certain threshold on the Jewish Names Index, and those who are not unique in the 1910 census and thus cannot be matched forward.

[24] 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.

Table 3 compares men who participated in the IRO program to other sets of household heads 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 in New York City and of 1896 for other Jewish households in New York City who lived outside of enclave neighborhoods. We thus control flexibly for year of arrival in the US in our analysis. This difference in average arrival year is partly 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.

At the time of removal, IRO participants had lower income scores than comparison enclave households (earning $723 in 1940 dollars, relative to $992 for other enclave residents in 1940 dollars). Jewish household who lived in more integrated New York neighborhoods, outside of enclaves, were considerably more affluent than their Lower East Side counterparts. Migration to these New York neighborhoods was likely out of the reach of the struggling immigrant families that the IRO sought to serve.

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 and country of birth, as well as all interactions.[25]

[25] This method follows Abramitzky, et al. (2020) and is similar to the machine-learning approach for computing income scores proposed by Twinam & Saavedra (2018). Note that the 1940 census

We also show results below using a modified income score that includes current state of residence in the prediction. 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).

Effect of IRO program participation on location

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."[26]

Table 3 describes the regional distribution of IRO participants based on their assignment location from 1899 to 1920 and their ultimate place of residence, as reported in the 1920 census.

---

does not record farm income. We compute income for farmers following Collins and Wanamaker (2017) 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.

[26] The IRO identified target locations through intermittent surveys and informal correspondence with established, but typically small, Jewish communities.

The majority of IRO participants were sent to towns and cities in the Midwest (64 percent), with approximately 20 percent being assigned to areas of the South and West. Only 15 percent of participants were resettled to Northeastern states. Internal IRO documentation reports that around 90 percent of participants were residing at the assignment location in the first year. By 1920, however, we find that only 15 percent of IRO households remain in the state to which they were assigned, and a large share were living in the Northeast again (68 percent).[27] We observe considerable variance in the "stickiness" of assignment locations: California and Minnesota retained 21 and 27 percent of their assignees, respectively, but only around 5 percent of assigned participants stayed in Indiana or Iowa.[28]

The internal correspondence and letters to the IRO underscore the unhappiness of many participants with their assignment locations (Rockaway, 1998).[29] Aaronson et al. (2020) find a similar pattern for participants in the Galveston Movement, a sister program of the IRO that redirected Jewish immigrants away from the Northeast and through the port of Galveston. Of the 10,000 Russia Jewish immigrants who arrived in Texas between 1907 and 1914, up to 90 percent

[27] For reference, 87 percent of our preferred comparison group – other Jews living in New York enclaves – still lived in the Northeast in 1920. We map these patterns in Appendix Figure 3.

[28] Twelve states account for 78 percent of assignment locations: Ohio, Missouri, Michigan, Illinois, Wisconsin, Pennsylvania, New York, Indiana, California, Minnesota, Nebraska, Iowa, Texas.

[29] The IRO archive contains many 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.*"

moved east of the Mississippi, mainly to the traditional Jewish enclaves in the Northeast and Midwest.

Despite the lack of stickiness of IRO participants in assignment locations, the IRO program did have a strong effect on the probability of leaving New York and moving out of Jewish enclaves. **Table 4** 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 11 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).

Figure 5 graphs the full distribution of neighborhood Jewish share for IRO and non-IRO participants before and after relocation. 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.

As with many mobility programs, the IRO program was a "bundled" treatment, shifting participants to neighborhoods with fewer co-ethnics *and* more higher status neighbors. Immigrant enclaves – both in New York City and other large metropolitan areas – were characterized not only by having a large foreign-born population but also by having residents of lower socio-economic status. **Appendix Table 1** documents that – not surprisingly – enumeration districts identified as "immigrant enclaves" had a higher immigrant share, but also had fewer homeowners and fewer residents working in white collar positions. **Figure 4** 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, English speaking share and homeownership rate.[30]

Estimation strategy

To study the association between residence in an immigrant enclave and economic assimilation, we compare the income score of IRO participants to neighboring residents of Jewish enclaves in New York City in the 1910 census, both before and after relocation. We stack data from two periods. 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:

$$y_{it} = \alpha_1 + \beta_1(IRO_i x\ After_t) + \beta_2(IRO_i) + \beta_3(After_t) + X_i\Gamma_1 + \varepsilon_{1,it} \qquad (1)$$

where the outcome variables *y* for household *i* include the logarithm of income score for fathers around 1910 and in 1920 or sons in 1940. The variable IRO_i is an indicator equal to one if an individual was ever part of the IRO program. The indicator $After$ is equal to one in 1920, by which point IRO participants will have been moved to new locations. 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 for comparison men 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

[30] Figure 4 is based on versions of equations (2) and (3) presented in the Estimation Strategy section below, each using an enumeration district characteristic as our outcome variables.

experienced greater earnings gains relative to non-participants after removal. If leaving Jewish enclaves in New York led to improved earnings, we expect $\beta_1 > 0$.[31]

The vector X_i includes fixed effects for a series of demographic and economic attributes interacted with the time period $After$ (=1920) to allow for differential trends by group. Most importantly, to allow for differential trends in earnings growth by initial economic characteristics, our preferred specification adds fixed effects for initial occupation and for placement in the initial income score distribution (in quintiles), along with interactions between these attributes and the $After_t$ indicator.[32] We also include interactions between individual year of birth, individual year of arrival in the US, and birth place (Russian/not) with the time period $After$.

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

$$Own\ name\ index_{i1910} = \alpha_2 + \beta_4 IRO_i + X_i \Gamma_2 + \varepsilon_{2,i} \quad (2)$$

[31] Our coefficient of interest β_1 is identical if we replace the IRO main effect with a set of individual fixed effects and only estimate the interaction between IRO and the $After$ indicator. We choose to show coefficients for the IRO main effect because it provides useful information about initial selection into the IRO program.

[32] 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 initial occupation and placement in the initial income score distribution are not identical because the income score is also based on age, state of residence, and country of birth.

We then assess whether men in the IRO program married less Jewishly-identified spouses, gave their children less Jewish-sounding names and learned English by 1920.

$$Wife/child\ name\ index_{i1920} = \alpha_3 + \beta_5 IRO_i + X_i \Gamma_3 + \varepsilon_{3,i} \quad (3)$$

$$Speaks\ English_{i1920} = \alpha_4 + \beta_6 IRO_i + X_i \Gamma_4 + \varepsilon_{4,i} \quad (4)$$

β_5 indicates whether program participants were more likely to marry wives with distinctively Jewish first names or to give their children Jewish names, and β_6 captures whether participants were more or less likely to speak English.[33] Equations (3) and (4) include a control for a man's own Jewish Names Index in the vector *X* to examine *changes* in cultural identity over time. If living outside of New York exposed participants to a wider range of cultural influences and expanded their pool of marriageable women, we expect $\beta_5 < 0$ and $\beta_6 > 0$; that is, we expect IRO participants to marry less Jewishly-identified spouses and to be more likely to speak English.

Occupational attainment and cultural assimilation after removal

We now turn to understanding the effect of leaving enclave neighborhoods on economic and cultural assimilation.

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). Column 3 stacks data from before and after program participation and reports coefficients from the difference-in-difference

[33] Note that, when we use *child name* as the dependent variable, the unit of observation is a child and when we use *wife name* as an outcome, we limit our sample to men who were not co-resident with a spouse at baseline. On average, we have two child-level observations per household head.

regression in equation (1). IRO participants start with a 21 percent earnings gap. Participating in IRO raises incomes by 23 percent, entirely erasing this initial gap.

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. Column 4 thus adds baseline controls for initial occupation and initial quintile in the income score distribution. After adding these controls, the baseline gap between IRO participants and non-participants falls to 2 percent, substantially balancing the initial gaps in economic activity. Participating in the IRO program raises income relative to this comparable group by 4.4 percent by 1920.

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 themselves were similarly negatively selected in this subsample – and completely converged by 1920. In the 1940 census, we find that the sons of IRO participants in 1940 earn 6 percent more than the sons of comparison households whose fathers held the same occupation at baseline, although the estimate is noisier because of our smaller sample size. Thus, the gains experienced by IRO households appear to be retained into the second generation.[34]

[34] **Appendix Table 4** reports results showing that sons of IRO participants attained slightly lower levels of schooling, perhaps because they were less able to take advantage of the investment in public colleges in New York City in the 1920s and 1930s.

Appendix Figure 2A more explicitly compares the intergenerational mobility of children of IRO participants and non-participants by initial rank in the national income distribution (Abramitzky et al., 2021c; Chetty et al., 2014; Chetty and Hendren, 2018).[35] We 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, indicating higher average levels of absolute mobility for any initial father rank. Moreover, the slope for IRO participants is substantially flatter than for non-participants, suggesting a weaker association between the initial rank of father and sons among IRO participants.[36] To put this in quantitative terms, the slope of the 1910 to 1920 binned income ranks is 0.36 for the comparison group and only 0.12 for IRO participants. 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. The program also appeared to have broader effects, however, in almost wiping the slate clean for IRO participants.

Table 6 explores further dimensions of the upward economic mobility experienced by IRO participants. Men who left New York City through IRO were no more likely to be in the labor force or to own a business that employed others. However, IRO participants were more likely to be self-employed and to work in professional or managerial roles. Correspondingly, they were less likely to work in the manufacturing sector, which was more prevalent in New York City than in

[35] In particular, we rank each son based on his income score relative to other sons born in the same year, and we rank fathers relative to all other fathers with sons born in the same year.

[36] **Appendix Figure 2B** produces a similar graph for intra-generational mobility, revealing a similar pattern relative to non-participants, particularly for those below the median.

other areas. Furthermore, IRO participants were more likely to be homeowners and more likely to have received citizenship by 1920.[37]

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. We consider a series of cultural outcomes in **Table 7**. First, we find that IRO participants were not selected on Jewish identity; rather, they held similarly Jewish-sounding names at baseline to comparison households (column 1). Second, IRO participants were two percentage points more likely to speak English by 1920 (column 2), perhaps because of their experience in neighborhoods and jobs where Yiddish was less common. Third, IRO participants married spouses with less Jewish-sounding names, an indication that exposure to life outside the enclave introduced them to a different pool of marriageable women (column 3). Yet, fourth, we find that IRO participants and their spouses select *more* Jewish-sounding names for their children, which is not consistent with the idea of changing cultural values (column 4). This pattern is similar for sons and daughters, but slightly larger and more statistically precise for sons. We will show below that these patterns are strikingly different by exposure to time spent out of New York.

The gains associated with leaving an enclave contrast with earlier findings for refugees resettled near others from their home country. We subdivide our sample into "likely refugees" and other economic migrants based on year of arrival and country of origin. Specifically, Jewish immigrants who left Russia between 1903 and 1906 are particularly likely to have been fleeing from anti-Jewish riots (pogroms), whereas other departures in our time period are less likely to be

[37] Catron (2019) has documented strong positive intergenerational effects of citizenship in this period.

refugee immigrants. **Table 8** subdivides IRO participants and comparison households into likely refugees and non-refugees. Only non-refugees appear to benefit from program participation (5 percent), suggesting that refugee migrants may depend more on the resources of ethnic enclaves. This pattern cautions against drawing wider lessons about the value of living in an immigrant neighborhood from a refugee sample alone.

Assimilation patterns by exposure to time outside of New York City enclaves

Men who voluntarily participated in the IRO program may have differed from their neighbors in unobservable ways. For example, men who were willing to leave the city may have been more resourceful or talented, even though they started out with lower income scores at baseline. Alternatively, some men sought out the IRO program following a spell of unemployment, and we may simply be capturing regression to the mean, akin to a classic Ashenfelter (1978) dip. One concern is that these personal attributes, rather than the mobility assistance through the IRO program, may help to explain the occupational attainment of program participants after removal. We address the possibility that IRO participants were selected on unobservable characteristics by considering differences *within* IRO participants who ended up with more or less exposure to life outside of a large Jewish enclave. By focusing on differences within IRO participants, we remove the initial selection bias that emerges when comparing program participants to non-participants.

For the program to improve economic outcomes, we assume that participants would need to leave the Jewish neighborhoods of New York for a non-trivial period of time. Indeed, 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. In particular, we compare men of the same age and arrival year in the US who moved through the IRO program in different years, generating different exposure to life

outside of New York City. We test whether removal year is associated with baseline attributes and do not find any selection into early/late removal.

In **Table 9**, we split IRO participants into three groups of roughly equal size based on when they were removed from the city: early removals (1900-06, 17 years in program on average), middle removals (1907-11, 11 years in program on average) and late removals (1912-1919, 5 years in program on average). The early and middle removal groups look similar in their initial income scores, with each earning 20 percent less than comparison households, while the late removal group was slightly less disadvantaged. Yet gains from the program monotonically increase with exposure to time outside of New York. IRO participants who were removed early earned 19 percent more than comparison households by 1920. In contrast, men who were removed in the middle of the program 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 close to a 1.0 percent gain in income score for each year spent out of the immigrant enclave.

Men who joined the IRO program earlier may have had fewer family connections in New York City, or more family outside of the city. We find little evidence for these alternative explanations. Early movers still enjoy the largest gains in **Appendix Table 5**, even after controlling for the number of likely Jews who shared an individual's surname (column 1) as a proxy for having family in New York, or directly adding surname fixed effects (column 2). The same pattern holds in column 3, 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). Furthermore, the benefits of early moves appear even when we drop men who joined the IRO program soon after arrival in the US; these men may have joined the program for different reasons – e.g., out of sense of adventure (column

4). 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.

Appendix Table 6 explores the effect of exposure to time outside of New York on cultural assimilation. Men who moved earlier are more likely to learn English. However, men who moved later in the program are more likely to marry a non-Jewish (or less Jewishly-identified) spouse. This pattern is more consistent with a change in marriageable pool, which could have been immediate, rather than with a shift in cultural attitudes, which would have taken some time and exposure to life outside of the enclave to occur.

Return migration to New York City

Nearly 50 percent of IRO participants moved back to New York City after some time spent away (see Table 2). We analyze who chose to return to New York and compare the outcomes of returners and non-returners, acknowledging that some component of this difference could be due to selection.

Table 10 starts by assessing two components of selection into return migration: initial income level and initial connection to Jewish culture as measured by Jewish Names Index value. IRO participants who chose to come back to New York were no different from non-returners on baseline income, but their names scored 1.8 points higher on the Jewish Names Index. Furthermore, return migrants gave their own children substantially more Jewish names after return (4.3 points).[38] This pattern suggests that one of the factors drawing migrants to return to New York City was the density of Jewish institutions and relationships available in enclave neighborhoods.

[38] 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 4).

IRO participants who did not return to New York by 1920 indeed 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 substantially change cultural attitudes.[39]

IRO participants who returned to New York ultimately ended up earning slightly more than participants who stayed outside (6 percent versus 3 percent). This advantage is entirely due to the higher wages in the New York metropolitan area. If we instead compare migrants on an income score that is location-invariant, we find that men who remained outside of New York fared better by moving farther up the occupational ladder (see **Figure 6** for details on this outcome).

Robustness

We make a number of decisions with our data in order to produce our main results. This section tests the robustness of our findings to each of these choices. In **Figure 6**, we present estimates from 18 separate robustness analyses of our main difference-in-difference estimate for the first generation from removal year to 1920 (shown above in Panel A, **Table 5**). These robustness estimates are derived from samples of various constructions and outcome variables.

We begin by examining differences based on our weighting decisions and the construction of our main comparison group. Our decision to weight the analysis sample to match the population in our main specification appears to have had no meaningful impact on our estimates. We could

[39] 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.

also have constructed our comparison group in several different ways. One option is to focus our comparison on non-participating households who had experienced unemployment at similar levels to the IRO participants. **Figure 6** shows an estimate based on comparing IRO to comparison households who had been unemployed for 12 or more weeks in 1909. This comparison produces consistent results. Alternatively, we might have defined our comparison group by whether they spoke Yiddish in the 1920 Census rather than by our Jewish index. Once again, this appears to produce no meaningful difference in our main estimate.

The decision to define the comparison group based on scoring above 1.4 on the Jewish index is also robust. Results are similar when we use a more stringent Jewish index threshold of 1.6 or 1.8, or when we include all men living in a Jewish enclave at baseline in our comparison group without requiring a "Jewish" name.

Our results are robust to our choice of matching algorithm. We present estimates from datasets constructed using three alternative algorithms: a modified version of the ABE algorithm that standardizes names using the NYSIIS phonetic algorithm (rather than using exact names as recorded); 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 (ABE Conservative); and a variant of the ABE algorithm that matches by first and last name, age and region of birth (rather than country of birth) to account for shifting borders in Eastern Europe over time (ABE Region).[40] In most

[40] We aggregated countries of birth into the regional coding scheme applied by IPUMS, where the birthplace codes (BPL) correspond to Northern Europe (400-419), Western Europe (420-429), Southern Europe (430-440), Central/Eastern Europe (450-459), and the Russian Empire (460-499). The small number of non-European birthplaces are grouped into an "Other" category.

cases, we continue to find an earnings gain of 3-5 percentage points relative to men who held the same occupation and income quintile at baseline.

The one noteworthy deviation from this pattern is the ABE Conservative algorithm. This particularly strict record linkage algorithm reduces the size of our IRO sample by almost 60 percent to approximately 900 observations. With this small sample, we do not have enough power to control for both occupation and income fixed effects. When we drop the 20 occupation fixed effects and control only for initial income quintile, results are similar to the main results.

We next make three cuts to the IRO sample: keep men who moved through the IRO to a preferred location; keep only men who stated no preferred location (known as "direct removals"); and drop men who report occupation strings like "no trade" in the IRO records for which there is no equivalent in the census data ("unusual occs"). Results are weaker when we split the sample by whether or not men stated a preferred location, but estimates are still marginally significant. Considering only men with common occupation strings raises the return to IRO participation (6 points).

Finally, we consider two alternative income scores. The first alternative a modified version of our "income score" that do not allow earnings to vary by current state of residence. The second is the standard 1950 "occupation score" or "occsore" that assigns each individual the median earnings for his occupation from the 1950 census. The IRO program gain is higher than six percentage points for alternate income score and over 3 percentage points for the occupation score.

Overall, we conclude that participating in the IRO program generated occupational income gains in all cases, with a consistent income gain of 3-4 percentage points for IRO program participants.

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 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 that leaving an enclave neighborhood need not come at a cost of losing cultural identity.

IRO participants who were exposed to more years outside of an enclave – either because they moved earlier in the program or because they remained out of New York by 1920 – experienced the largest gains in income score. In contrast, men who chose to move back to New 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 for refugee migrants 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 – we find lower gains for leaving enclaves for Jewish migrants who were likely fleeing from persecution. Understanding when and who are helped by enclave neighborhoods is a fruitful avenue for future research.

References

Aaronson, D., Davis, J., Schulze, K., 2020. Internal immigrant mobility in the early 20th century: evidence from Galveston, Texas. Explor. Econ. Hist. 76, 101317.

Abramitzky, R., Boustan, L., 2017. Immigration in American economic history. J. Econ. Lit. 55, 1311–45.

Abramitzky, R., Boustan, L., Catron, P., Connor, D., Voigt, R., 2021a. Refugees without Assistance: English-Language Attainment and Economic Outcomes in the Early Twentieth Century.

Abramitzky, R., Boustan, L., Eriksson, K., 2020. Do immigrants assimilate more slowly today than in the past? Am. Econ. Rev. Insights 2, 125–41.

Abramitzky, R., Boustan, L., Eriksson, K., 2019. To the new world and back again: Return migrants in the age of mass migration. ILR Rev. 72, 300–322.

Abramitzky, R., Boustan, L., Eriksson, K., 2016. Cultural Assimilation during the Age of Mass Migration. NBER Work. Pap. Ser.

Abramitzky, R., Boustan, L., Eriksson, K., 2014. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. J. Polit. Econ. 122.

Abramitzky, R., Boustan, L., Eriksson, K., 2012. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. Am. Econ. Rev. 102, 1832–1856.

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., Pérez, S., 2021b. Automated linking of historical data. J. Econ. Lit. 59, 865–918.

Abramitzky, R., Boustan, L., Jacome, E., Perez, S., 2021c. Intergenerational Mobility of Immigrants in the United States over Two Centuries. Am. Econ. Rev. 111, 580–608.

Akbar, P.A., Hickly, S.L., Shertzer, A., Walsh, R.P., 2022. Racial segregation in housing markets and the erosion of black wealth. Rev. Econ. Stat. 1–45.

Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. Rev. Econ. Stat. 47–57.

Bazzi, S., Gaduh, A., Rothenberg, A.D., Wong, M., 2019. Unity in diversity? How intergroup contact can foster nation building. Am. Econ. Rev. 109, 3978–4025.

Beaman, L.A., 2012. Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US. Rev. Econ. Stud. 79, 128–161.

Borjas, G.J., 2000. Ethnic enclaves and assimilation. Swed. Econ. Policy Rev. 7, 89–122.

Boustan, L., 2007. Were Jews political refugees or economic migrants? Assessing the persecution theory of Jewish emigration, 1881–1914. New Comp. Econ. Hist. Essays Honor Jeffrey G Williamson MIT Press Camb.

Bryan, G., Chowdhury, S., Mobarak, A.M., 2014. Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. Econometrica 82, 1671–1748.

Card, D., 1996. The effect of unions on the structure of wages: A longitudinal analysis. Econom. J. Econom. Soc. 957–979.

Catron, P., 2019. The citizenship advantage: Immigrant socioeconomic attainment in the age of mass migration. Am. J. Sociol. 124, 999–1042.

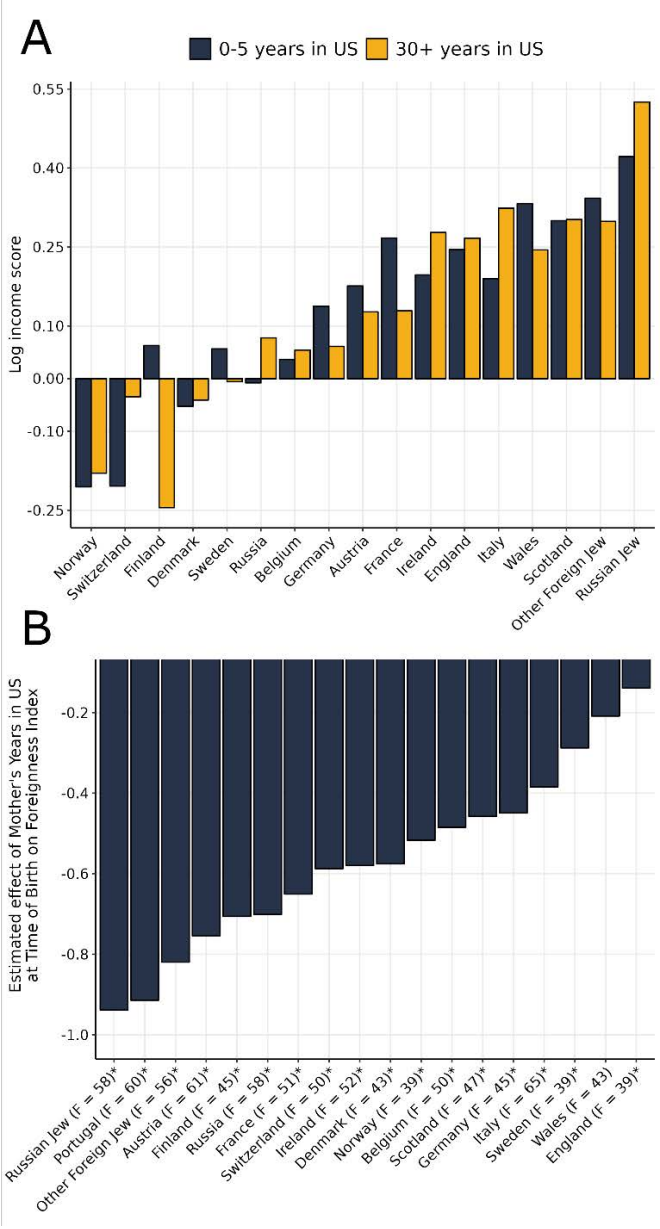
Chetty, R., Hendren, N., 2018. The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. Q. J. Econ. 133, 1107–1162.

Chetty, R., Hendren, N., Kline, P., Saez, E., Turner, N., 2014. Is the United States still a land of opportunity? Recent trends in intergenerational mobility. Am. Econ. Rev. 104, 141–147.

Chiswick, B.R., 1999. The occupational attainment and earnings of American Jewry, 1890 to 1990. Contemp. Jew. 20, 68–98.
Chiswick, B.R., 1992. Jewish immigrant wages in America in 1909: An analysis of the Dillingham Commission data. Explor. Econ. Hist. 29, 274–289.
Chiswick, B.R., 1983. The earnings and human capital of American Jews. J. Hum. Resour. 313–336.
Collins, W.J., Zimran, A., 2019. The economic assimilation of Irish Famine migrants to the United States. Explor. Econ. Hist. 74, 101302.
Connor, D.S., 2020. Class Background, Reception Context, and Intergenerational Mobility: A Record Linkage and Surname Analysis of the Children of Irish Immigrants. Int. Migr. Rev. 54, 4–34.
Connor, D.S., 2019. The Cream of the Crop? Geography, networks and Irish migrant selection in the Age of Mass Migration. J. Econ. Hist. 79.
Connor, D.S., Gutmann, M.P., Cunningham, A.R., Clement, K.K., Leyk, S., 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. Ann. Am. Assoc. Geogr. 1–18.
Connor, D.S., Storper, M., 2020. The changing geography of social mobility in the United States. Proc. Natl. Acad. Sci. 117, 30309–30317.
Damm, A.P., 2009. Ethnic enclaves and immigrant labor market outcomes: Quasi-Experimental Evidence. J. Labor Econ. 27, 281–314.
Danzer, A.M., Yaman, F., 2013. Do ethnic enclaves impede immigrants' integration? Evidence from a quasi-experimental social-interaction approach. Rev. Int. Econ. 21, 311–325.
Dillingham, W.P., 1911. Reports of the Immigration Commission. US Government Printing Office.
Diner, H.R., 2000. Embracing" World of Our Fathers": The Context of Reception. Am. Jew. Hist. 88, 449–462.
Edin, P.-A., Fredriksson, P., Aslund, O., 2003. Ethnic Enclaves and the Economic Success of Immigrants--Evidence from a Natural Experiment. Q. J. Econ. 118, 329–357. https://doi.org/10.1162/00335530360535225
Eisenberg, E., 1995. Jewish Agricultural Colonies in New Jersey: 1882-1920. Syracuse University Press.
Ellis, M., Wright, R., Parks, V., 2004. Work together, live apart? Geographies of racial and ethnic segregation at home and at work. Ann. Assoc. Am. Geogr. 94, 620–637.
Eriksson, K., 2020. Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration. Eur. Rev. Econ. Hist. 24, 427–446.
Eriksson, K., Ward, Z., 2019. The residential segregation of immigrants in the United States from 1850 to 1940. J. Econ. Hist. 79, 989–1026.
Fridkis, A.L., 1981. Desertion in the American Jewish immigrant family: The work of the National Desertion Bureau in cooperation with the Industrial Removal Office. Am. Jew. Hist. 71, 285–299.
Glazier, J., 1998. Dispersing the Ghetto: The Relocation of Jewish Immigrants Across America. Cornell University Press.
Industrial Removal Office, 1911. Tenth Annual Report of the Industrial Removal Office for the Year 1910. New York.

Kahan, A., 1978. Economic Opportunities and Some Pilgrims' Progress: Jewish Immigrants from Eastern Europe in the U.S., 1890–1914. J. Econ. Hist. 38, 235–251. https://doi.org/10.1017/S0022050700088276
Kuznets, S., 1975. Immigration of Russian Jews to the United States: background and structure. Charles Warren Center for Studies in American History.
Laliberté, J.-W., 2019. Language skill acquisition in immigrant social networks: Evidence from Australia. Labour Econ. 57, 35–45.
Lieberson, S., 1980. A piece of the pie: Blacks and white immigrants since 1880. Univ of California Press.
Lodge, H.C., 1909. Speeches and addresses, 1884-1909. Houghton Mifflin.
Maffi, M., 1994. Gateway to the promised land: ethnic cultures on New York's Lower East Side. Rodopi.
Marinbach, B., 2012. Galveston: Ellis Island of the West. SUNY Press.
Pagnini, D.L., Morgan, S.P., 1990. Intermarriage and social distance among US immigrants at the turn of the century. Am. J. Sociol. 96, 405–432.
Riis, J.A., 1890. How the other half lives: studies among the tenements of New York. Charles Scribner's Sons, New York.
Rockaway, R.A., 1998. Words of the uprooted: Jewish immigrants in early twentieth-century America. Cornell University Press.
Rosenthal, E., 1975. The equivalence of United States census data for persons of Russian stock or descent with American Jews: An evaluation. Demography 12, 275–290.
Rosenwaike, I., 1971. The Utilization of Census Mother Tongue Data in American Jewish Population Analyses. Jew. Soc. Stud. 33, 141–159.
Shertzer, A., Walsh, R.P., 2019. Racial sorting and the emergence of segregation in American cities. Rev. Econ. Stat. 101, 415–427.
Spitzer, Y., 2021. Pogroms, networks, and migration: The Jewish migration from the Russian Empire to the United States 1881–1914. Maurice Falk Inst. Econ. Res. Isr. Discuss. Pap. Ser. 0_1-83.
Szajkowski, Z., 1973. The" Yahudi" and the Immigrant: A Reappraisal. Am. Jew. Hist. Q. 63, 13–44.
Twinam, T., Saavedra, M., 2018. A Machine Learning Approach to Improving Occupational Income Scores.
Waldinger, R., 1986. Through the Eye of the Needle: Immigrants and Enterprise in. New York Univeristy Press, New York.
Ward, Z., 2020. The low return to English fluency during the Age of Mass Migration. Eur. Rev. Econ. Hist. 24, 219–242.
Xie, Y., Gough, M., 2011. Ethnic enclaves and the earnings of immigrants. Demography 48, 1293–1315.
Xu, D., 2020. The effects of immigration restriction laws on immigrant segregation in the early twentieth century US. J. Comp. Econ. 48, 422–447.
Zipperstein, S.J., 2018. Pogrom: Kishinev and the tilt of history. Liveright Publishing.

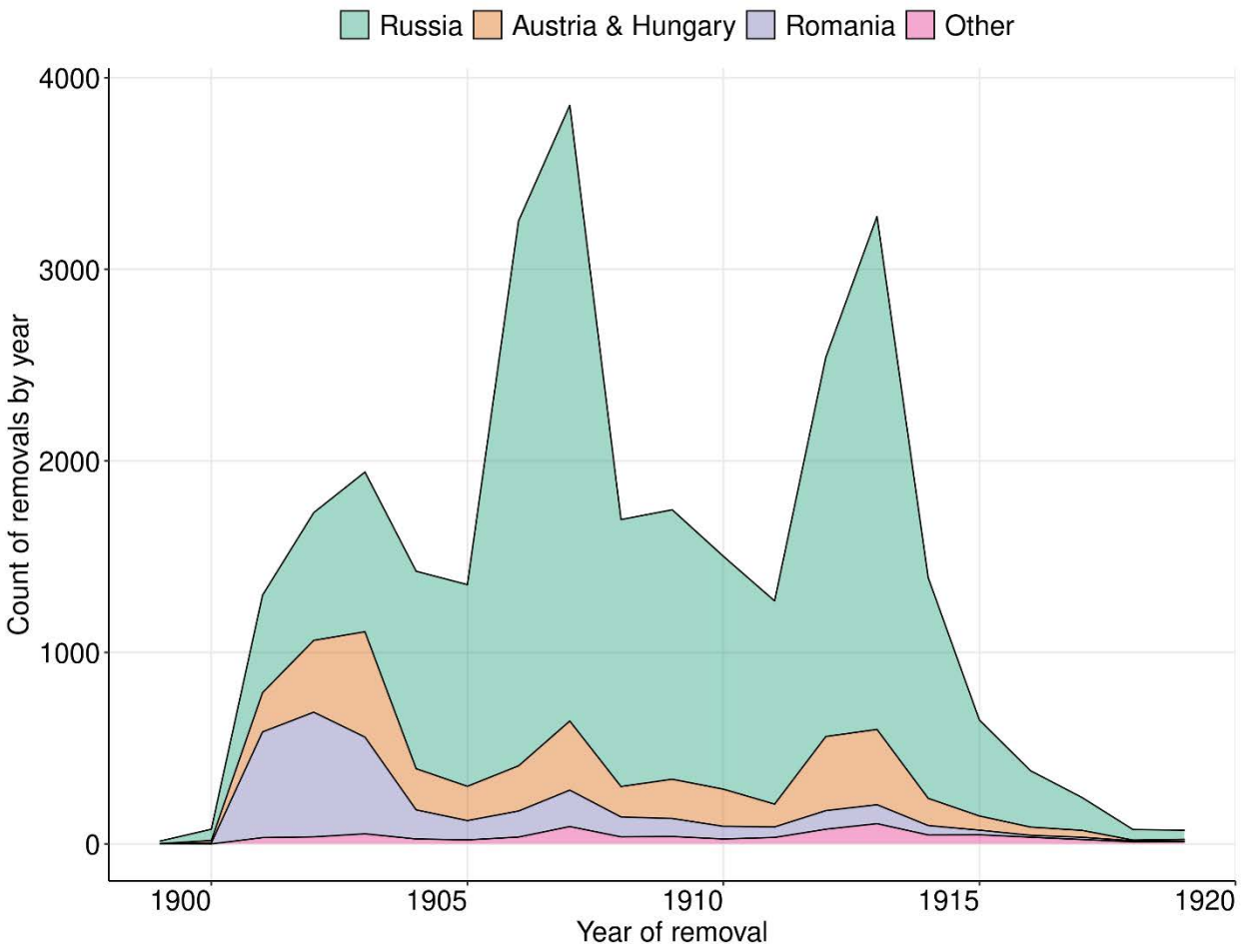
Figure 1: Log income score of immigrants by country of origin relative to the US-born, and changes in children's Foreignness Name Index for each year immigrant mother spent in the US



Notes: Two figures showing indicators of economic and cultural assimilation among European immigrants. Panel A shows the 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. Panel B shows the association between mother spending additional year in the US at time of child's birth and the foreignness index of a child's name, by ethnicity or country of origin.

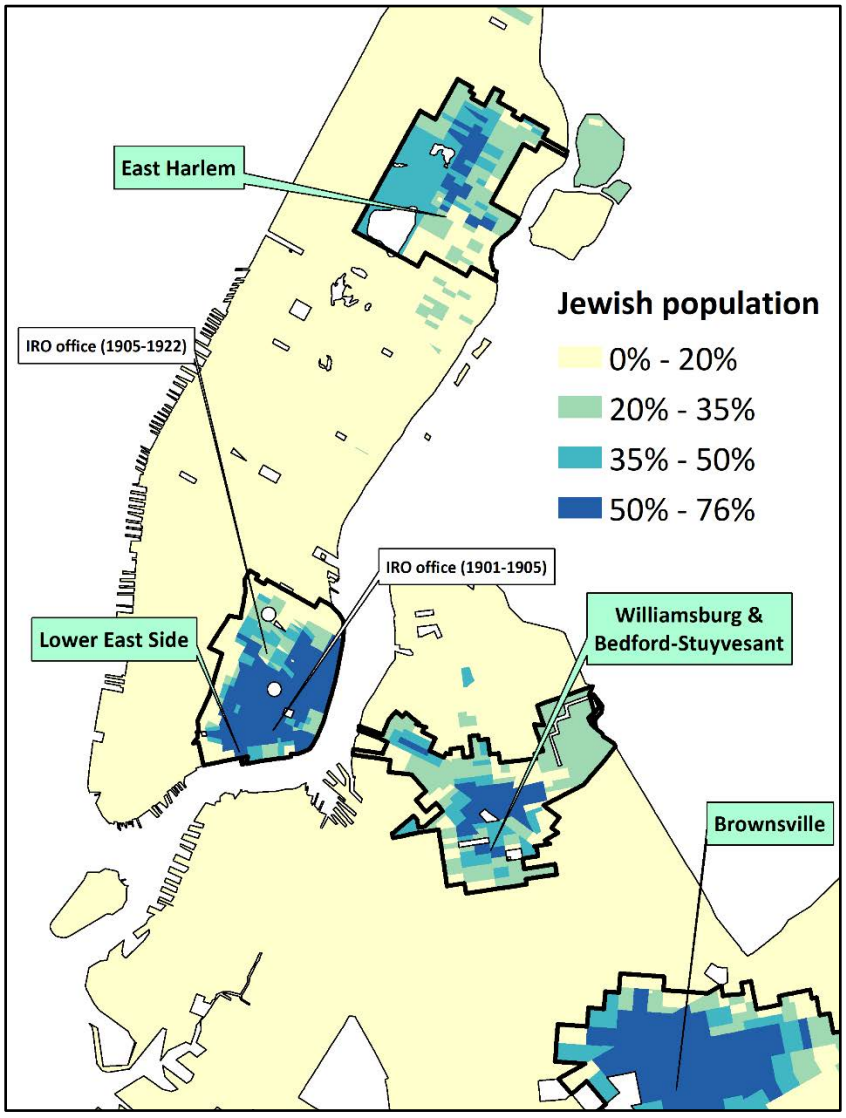
Estimates come from a regression of the foreignness index of a child's name on a set of interactions between mother's country of birth or ethnicity and years mother spent in the US at time of birth. This graph is a replication of Figure 2 (Panel A) from Abramitzky, Boustan and Eriksson, 2020 that distinguishes 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 2: Frequency of Industrial Removal Office resettlements by birthplace, 1899-1919



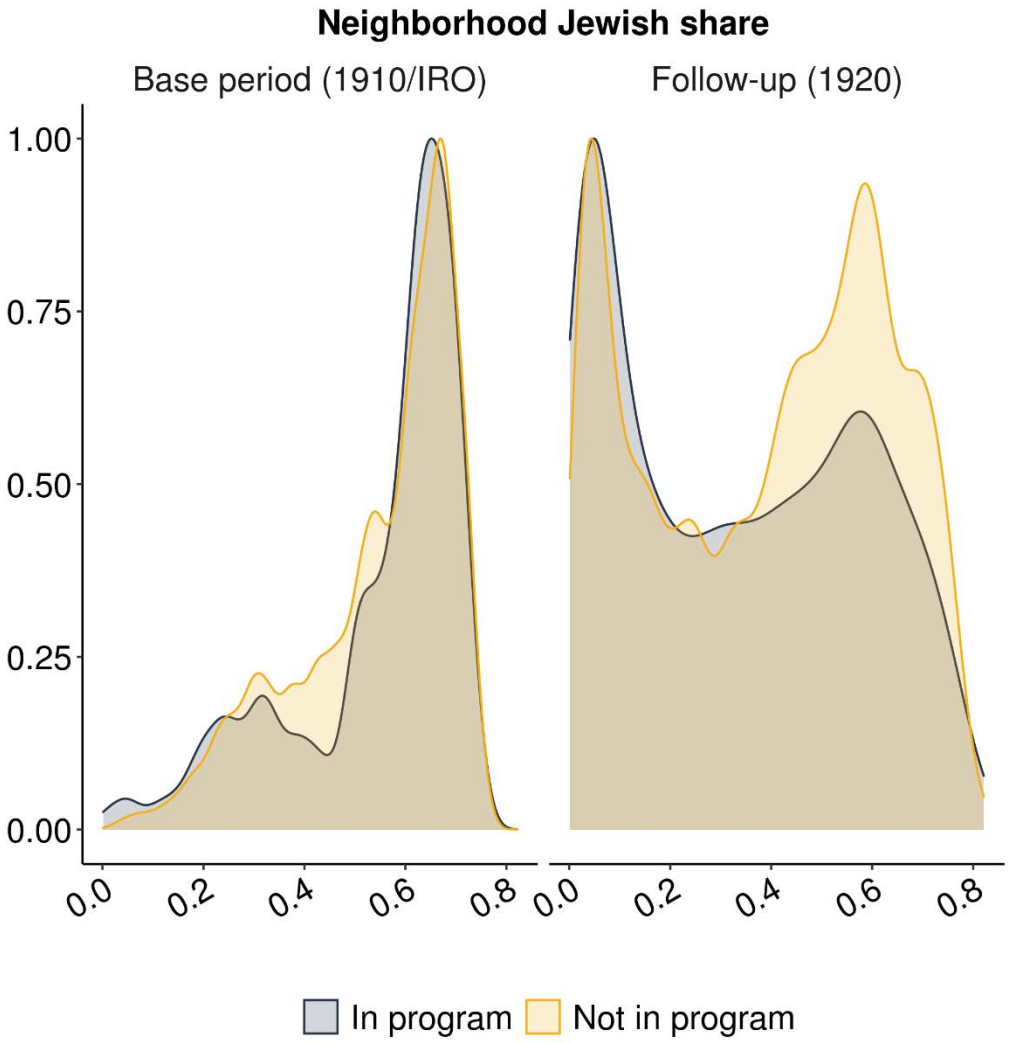
Notes: Yearly frequency of IRO resettlements between 1899 and 1920 by country of birth, based on the IRO record books.

Figure 3: Delineation of New York Jewish enclave boundaries by the Jewish share of enumeration districts in 1910



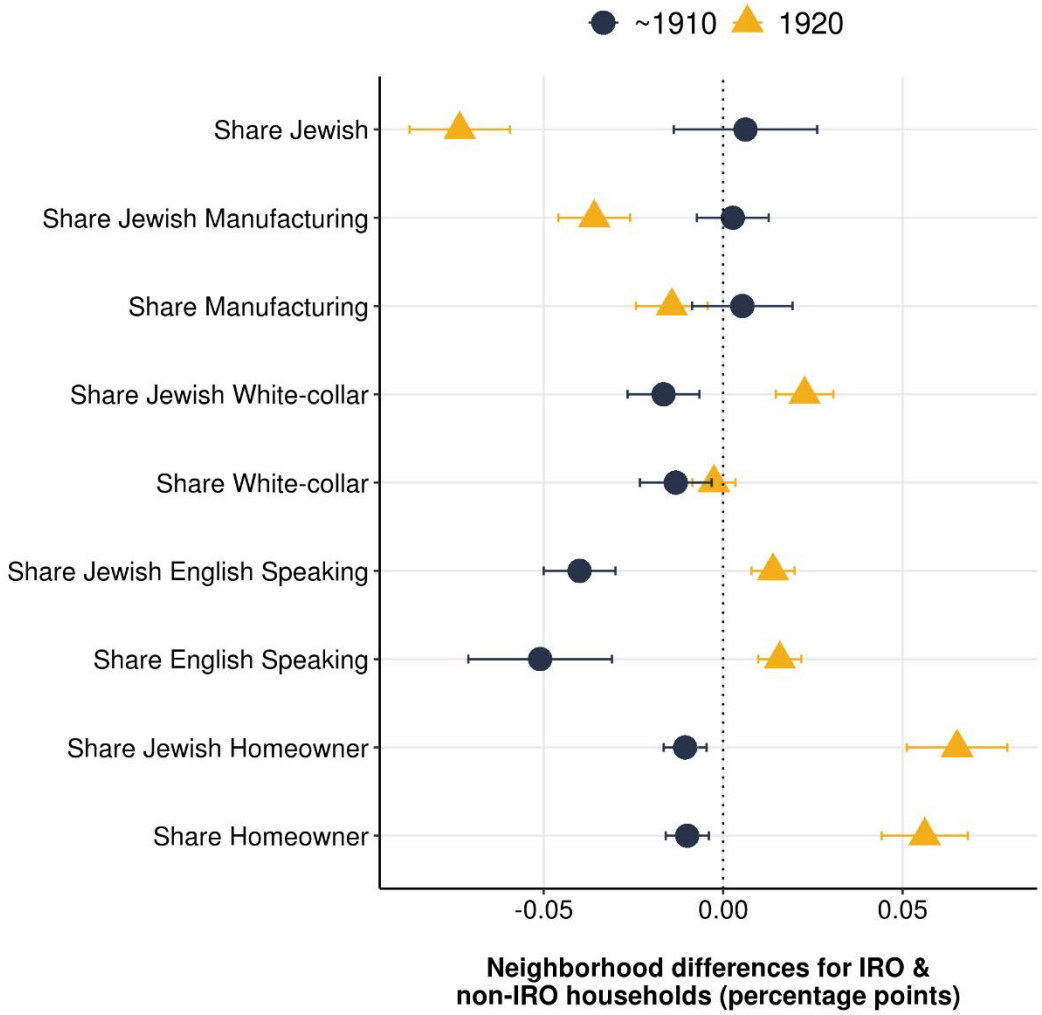
Notes: The boundaries of Jewish enclaves in New York superimposed on 1910 enumeration district boundaries. Boundaries of enclaves are determined by the Jewish population share of enumeration districts. The Jewish population share of enumeration districts is calculated from the share of individuals with 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 4: Jewish share of neighborhoods between IRO participants and Jewish households in New York enclaves circa 1910



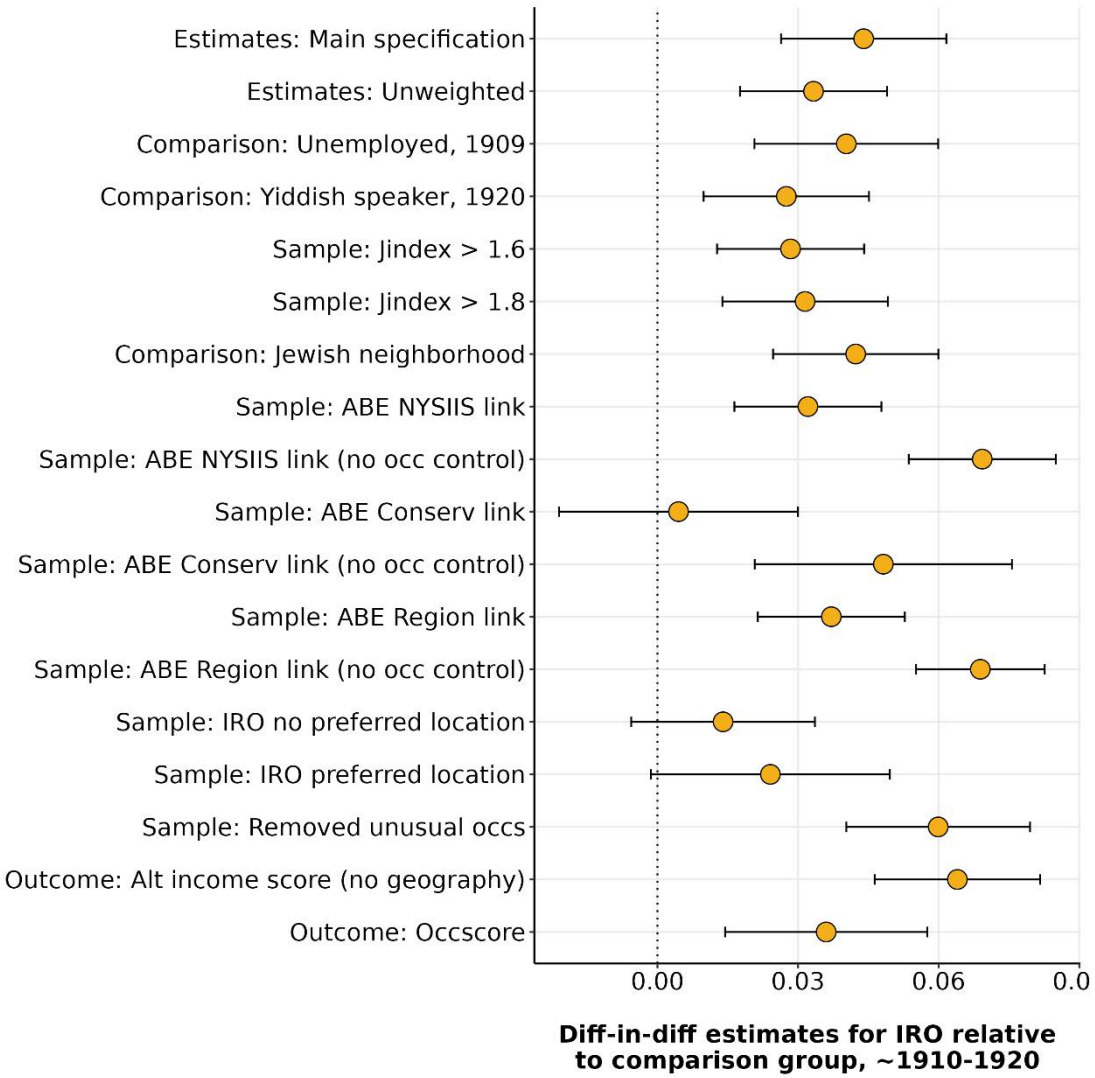
Notes: Kernel density plot of Jewish share of enumeration district circa 1910 and in 1920 for IRO and other Jewish households living in New York enclaves at baseline.

Figure 5: 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). Standard errors clustered by 1910 ED of residence.

Figure 6: Comparing IRO participants to residents of New York enclaves across various robustness specifications, ~1910-1920



Notes: A figure showing difference-in-difference estimates from 18 separate models with varying sample restrictions and outcomes. The estimates are derived from models identical to that shown in the main specification (Panel A, Column 3, Table 5). We show the main specification (1) and estimates from a sample that is not reweighted for linkage (2). We try defining the comparison group based on being unemployed for 12 or more weeks in 1909 (3), based having a Yiddish mother tongue in the 1920 census (4), by a more stringent criteria on the Jewish index (5-6), or with no mother tongue restriction but based on living in a Jewish neighborhood (7). We link all observations based on various alternative common linkage approaches (8-13). We constrain the IRO sample by whether or not they specified a preferred assignment location (14-15), dropping individuals that reported "no trade" at baseline (16), and testing against two alternative outcomes: an income score measure that does not allow for geographic various in earnings by occupation (17) or by the standard occupational score (18).

Tables

Table 1: 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	1.98
	2	HYMAN	GOLDBERG	257	1.98
	3	HYMAN	GOLDSTEIN	229	1.98
	4	HYMAN	COHEN	687	1.98
	5	MEYER	COHEN	334	1.98
	6	ISIDORE	COHEN	236	1.98
	7	ISRAEL	COHEN	203	1.98
	8	ABRAHAM	SHAPIRO	245	1.98
	9	ABRAHAM	KAPLAN	265	1.97
	10	ABRAHAM	LEVINE	435	1.97
Borderline	463	BENJAMIN	HARRIS	418	1.42
	464	HERMAN	SCHULTZ	614	1.42
	465	HARRY	SHAFFER	352	1.42
	466	ALEX	MILLER	291	1.41
	467	JOSEPH	WERNER	222	1.41
	468	SAMUEL	TUCKER	249	1.41
Least Jewish	17,426	CLARENCE	BOYD	223	0.03
	17,427	JUAN	MARTINEZ	656	0.03
	17,428	JUAN	RODRIGUEZ	256	0.03
	17,429	FRANCISCO	MARTINEZ	370	0.02
	17,430	BOOKER	WASHINGTON	247	0.02
	17,431	CLYDE	COX	205	0.02
	17,432	FLOYD	COX	230	0.02
	17,433	CLYDE	CAMPBELL	285	0.01
	17,434	FLOYD	CAMPBELL	232	0.01
	17,435	WADE	HAMPTON	217	0.01

Notes: Jewishness of a selection of the 33,661 names in the 1920 census held by at least 200 males. 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 2: 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	0.62
Mean (and st. dev.) age at removal	28 (8.79)
Mean (and st. dev.) arrival year in US	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
Region (assigned / resident 1920)	
Northeast	0.15 / 0.68
Midwest	0.64 / 0.22
South	0.12 / 0.05
West	0.09 / 0.05

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). Our linked sample (N = 3,612) includes men with complete information on name and age whose names are above 1.4 on the Jewish Names Index and who are unique by name and age in the 1910 census. We report summary statistics for observations that have complete information on occupation and neighborhood (N = 2,362). Income scores are based on imputation from 1940 census.

Table 3: Summary statistics for IRO participants and various comparison groups

		Foreign-born, likely Jews		
	IRO	Lived in NYC in enclave, 1910	Lived in NYC outside enclave, 1910	Lived outside NYC, 1910
Demographic and economic				
Age, 1920	38	40	43	42
Arrival year	1903	1900	1896	1897
Income score, ~1910	$723	$992	$1234	$992
Income score, 1920	$1270	$1315	$1427	$1254
Second gen. income score, 1940	$1257	$1348	$1402	$1223
New York resident, 1920	0.46	0.72	0.74	0.18
Lives in assigned state, 1920	0.15	-	-	-
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
Speaks English, 1920	0.95	0.94	0.96	0.94
Observations (*N*)	1,486	5,416	1,078	5,962

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. 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. The dependent variable from our main specifications are based on the natural log of the income scores presented in the table above.

Table 4: Outmigration rate by 1920 for IRO program participants and comparison sample

	Outmigration rate by 1920	
	New York City	Jewish Enclave
IRO	53%	57%
Not in IRO	27%	46%

Notes: Outmigration rate for IRO participants relative to other Jewish men in comparison sample, as defined in text. Outmigration measured as living outside of the New York City state economic area in 1920 (left) or living outside of a Jewish enclave in 1920 (right). Enclaves are defined as enumeration districts anywhere in the United States that were at least 40 percent Jewish in 1920.

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)
	~1910	1920	~1910-1920	~1910-1920
A. First generation				
IRO	-0.180***	-0.0192*	-0.212***	-0.0224***
	(0.007)	(0.008)	(0.007)	(0.003)
In 1920			0.922***	1.478***
			(0.023)	(0.039)
IRO *x* In 1920			0.226***	0.0440***
			(0.011)	(0.009)
N	22108	22108	44216	44216
	~1910	1940	~1910-1940	~1910-1940
B. Second generation				
IRO	-0.185***	0.0307	-0.185***	-0.0103
	(0.015)	(0.041)	(0.013)	(0.006)
In 1940			-4.945***	-4.649***
			(0.399)	(0.472)
IRO *x* In 1940			0.216***	0.0633
			(0.040)	(0.044)
N	4554	4554	9108	9108
Controls				
Birth cohort	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y
~1910 Occ.	N	N	N	Y
~1910 Inc. rank	N	N	N	Y

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

Notes: Log income score difference between IRO and other Jews living in New York enclaves in 1910. Reference category are Jews living in New York enclaves in 1910. Observations are restricted to have a Jewish index > 1.4, foreign-born, aged 26-59 in 1920 and have a reported occupation in the base period and in 1920. The difference-in-difference coefficients (Column 3) are estimated from an interaction between IRO and a dummy variable based on period of observation (post-1920 for first generation, post-1940 for second generation). Controls in the diff-in-diff models are estimated with a main effect and an interaction with the period dummy. Linear term for age at first observation included as additional continual control variable for IRO. For the second-generation sons, aged 18 to 41 in 1940, the dependent variable is the log of actual income

in 1940 dollars). Observations are reweighted by their probability of selection into sample through record linkage (see Data Appendix). Standard errors clustered at household for second generation.

Table 6: Other economic outcomes for IRO participants, 1920

	In labor force	Employer	Self-employed	Professional worker	Manufact worker	Citizen	Owns home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1920	1920	1920	1920	1920	1920	1920
IRO	-0.0012	-0.00266	0.0300**	0.0372***	-0.0269*	0.0277*	0.0251*
	(0.001)	(0.007)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
N	22108	22108	22108	22108	22108	22108	22108
Mean of dependent var, comparison group	0.99	0.11	0.30	0.34	0.27	0.56	0.17
Controls							
Birth cohort	Y	Y	Y	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y	Y	Y	Y
Russian	Y	Y	Y	Y	Y	Y	Y
~1910 Inc score	Y	Y	Y	Y	Y	Y	Y

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

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

Table 7: Cultural assimilation of IRO participants, 1920

	Own Jewish index	Speaks English	Wife's Jewish index	Jewish index of children		
				All	*Sons*	*Daughters*
	(1)	(2)	(3)	(4)	(5)	(6)
	~1910	1920	1920	1920	1920	1920
IRO	0.00776	0.0204*	-0.0180**	0.0185*	0.0253*	0.0156
	(0.005)	(0.008)	(0.009)	(0.008)	(0.012)	(0.011)
N	6883	6883	6883	12300	6306	5994
Controls						
Birth cohort	Y	Y	Y	Y	Y	Y
Arrival Year	Y	Y	Y	Y	Y	Y
Russian birthplace	Y	Y	Y	Y	Y	Y
Own Jewish index	N	Y	Y	Y	Y	Y
English speaking HH	Y	Y	Y	N	N	N
Child: age, sex, foreign	N	N	N	Y	Y	Y
Household clustered SEs	N	N	N	Y	Y	Y

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

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

Table 8. Log income score of IRO participants in 1920, by refugee status

	Cross-section		Diff-in-diff
	(1)	(2)	(3)
	~1910	1920	~1910-1920
Reference = Not IRO, not refugee			
IRO, not refugee	-0.171***	0.00398	0.0636***
	(0.013)	(0.020)	(0.017)
IRO, refugee	-0.203***	-0.0404**	0.00857
	(0.012)	(0.013)	(0.014)
Not IRO, refugee	0.000555	0.0161	0.00867
	(0.010)	(0.015)	(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

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

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

Table 9: 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	0.188***
	(0.027)	(0.027)	(0.016)
IRO: 8-13 years (middle)	-0.218***	-0.0135	0.0573***
	(0.011)	(0.012)	(0.012)
IRO: 1-7 years (late)	-0.147***	-0.0217	-0.0896***
	(0.017)	(0.017)	(0.013)
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

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

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

Table 10: Economic and cultural assimilation of IRO participants by return to New York

	Income score (cross-section)		Income score (diff-in-diff)	
	(1)	(2)	(3)	(4)
	~1910	1920	~1910-1920	~1910-1920
A. Compliance with relocation				
IRO: Returned to NYC	-0.175***	0.00205	0.224***	0.0558***
	(0.009)	(0.010)	(0.014)	(0.011)
IRO: Stayed outside NYC	-0.184***	-0.0391***	0.228***	0.0335**
	(0.009)	(0.011)	(0.014)	(0.012)
N	22108	22108	44216	44216
	Own Jewish index	English speaking HH	Wife's Jewish index	Child's Jewish index
	~1910	1920	1920	1920
B. Compliance with relocation				
IRO: Returned to NYC	0.0179***	-0.00273	0.00196	0.0426***
	(0.006)	(0.012)	(0.011)	(0.010)
IRO: Stayed outside NYC	-0.00153	0.0417***	-0.0362***	-0.00109
	(0.007)	(0.009)	(0.011)	(0.010)
N	6883	6883	6883	12300
Controls				
Cohort, arrival year, birthplace	Y	Y	Y	Y
~1910 Occ [a]	N	N	N	Y
~1910 Inc. rank [a]	N	N	N	Y
Own Jewish index [b]	N	N	Y	Y
Child: age, sex, foreign born [b]	N	N	N	Y

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

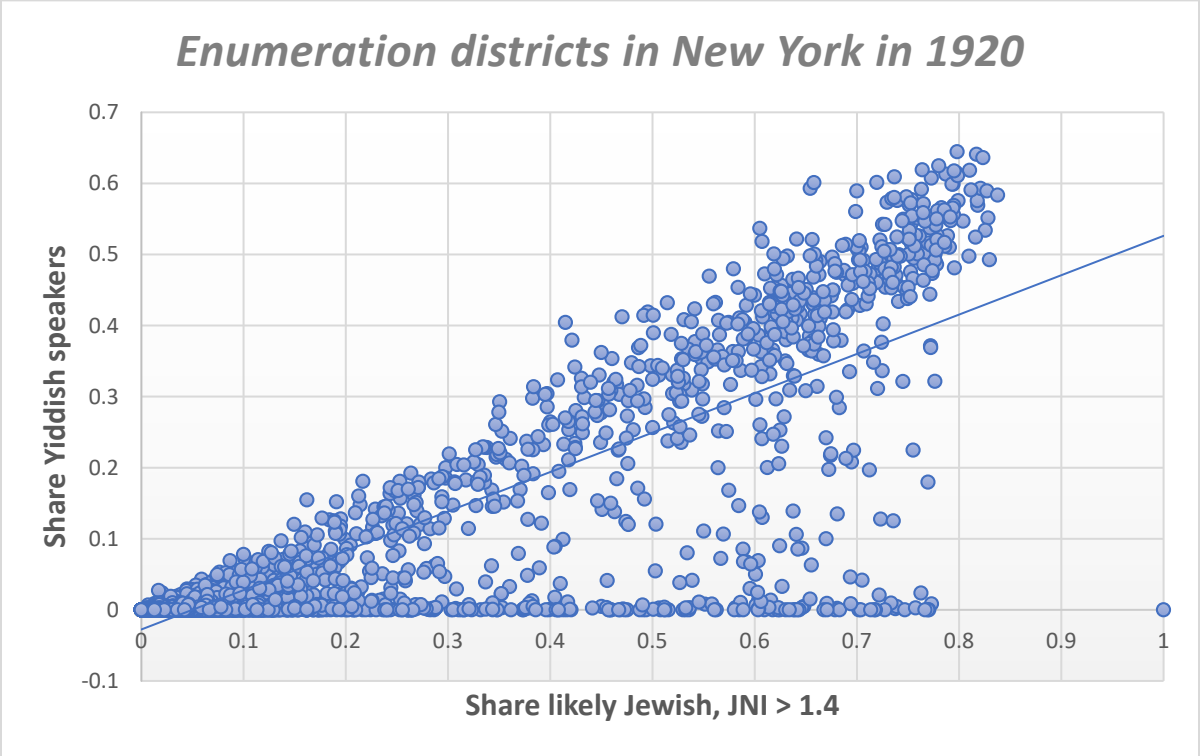
Notes: Economic and cultural assimilation by participants decision to return to New York. Reference category are Jews living in New York enclaves in 1910. The first-generation sample is restricted to household heads in 1920 who were not co-resident with a spouse in the base period, and with a Jewish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

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

June 2023

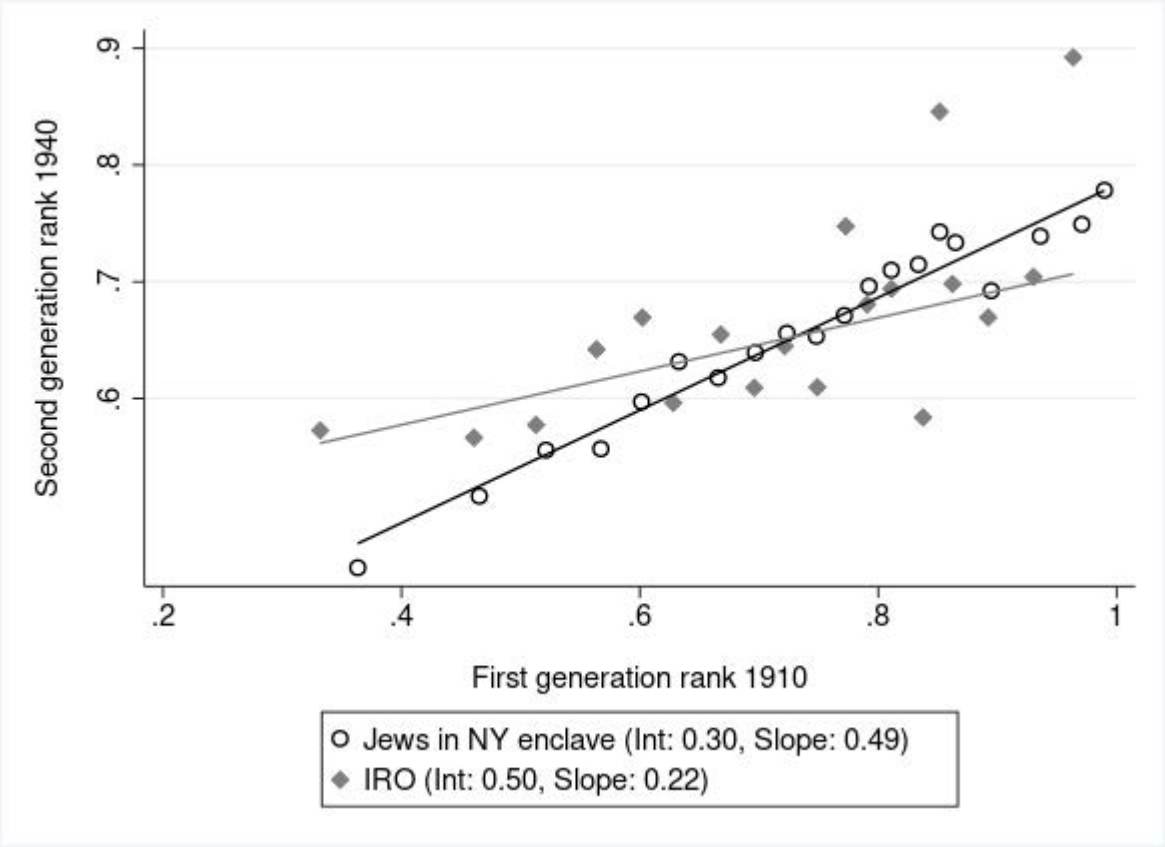
Appendix -- For On-line Publication Only

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



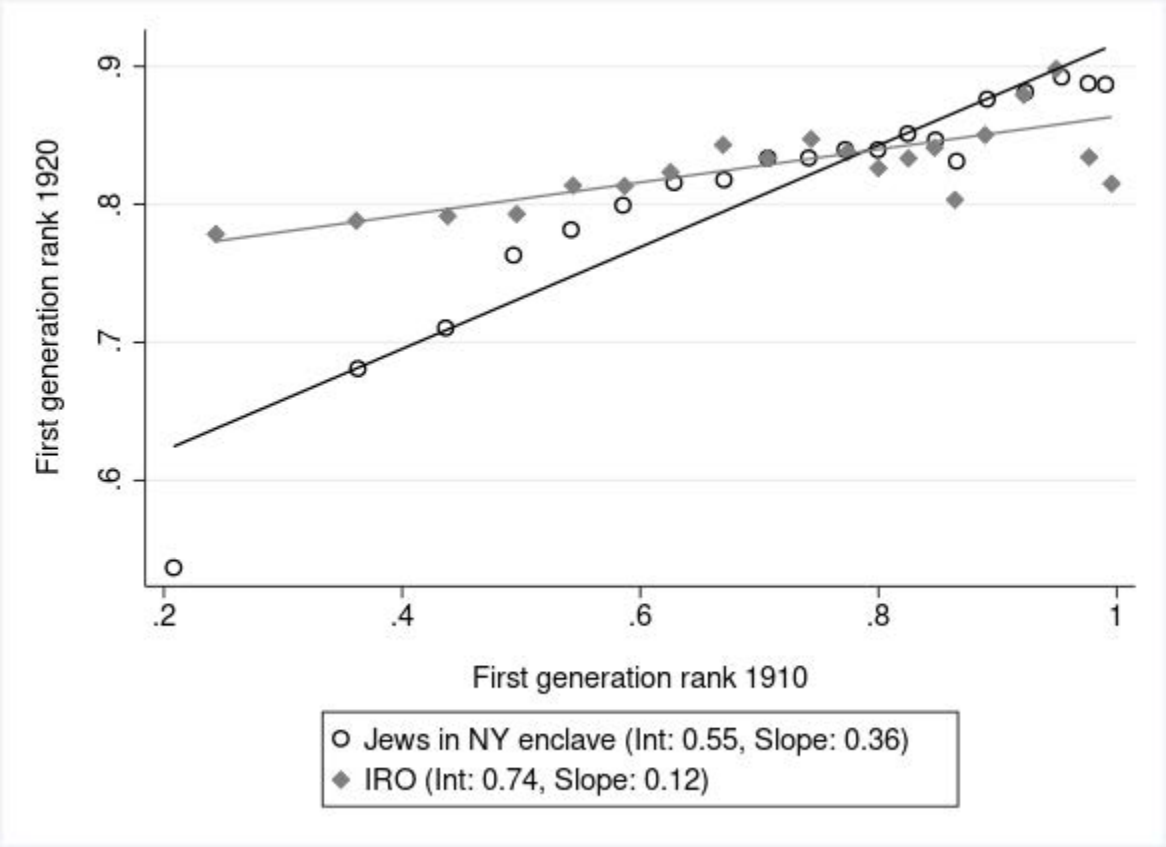
Notes: A figure showing the relationship between the share of Yiddish speakers and the share of likely Jews in New York enumeration districts in 1920. The correlation between these ED-level measures in New York City is +0.85.

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



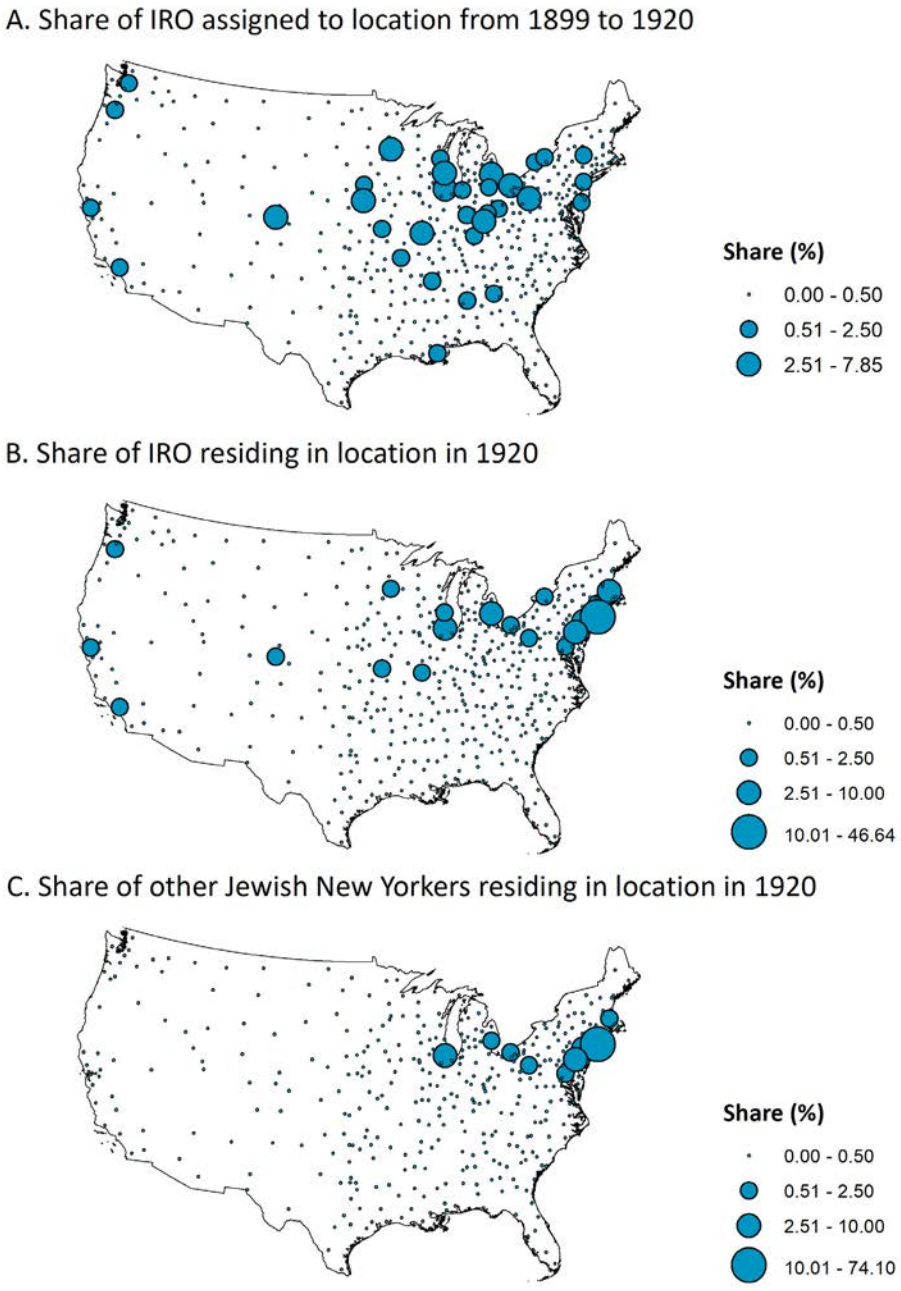
Notes: Binned scatterplot graphing the 1910 income rank of IRO participants and those of their sons in 1940 against the corresponding values for other Jewish immigrants in New York enclaves (circa 1910). The first and second generation in each group are assigned percentile ranks based on their log income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

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



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

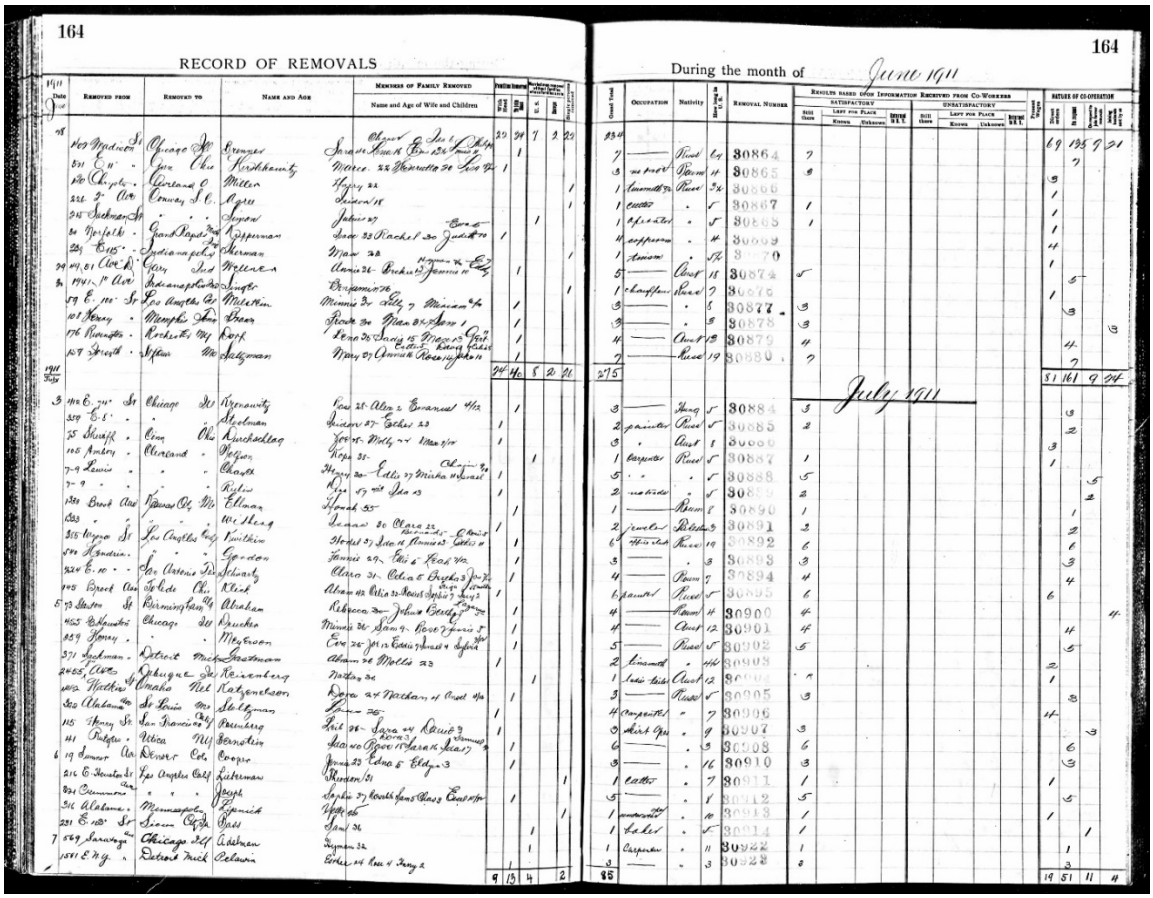
Appendix Figure 3: Distribution of IRO and other foreign-born Jewish households across state economic areas, 1899-1920



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

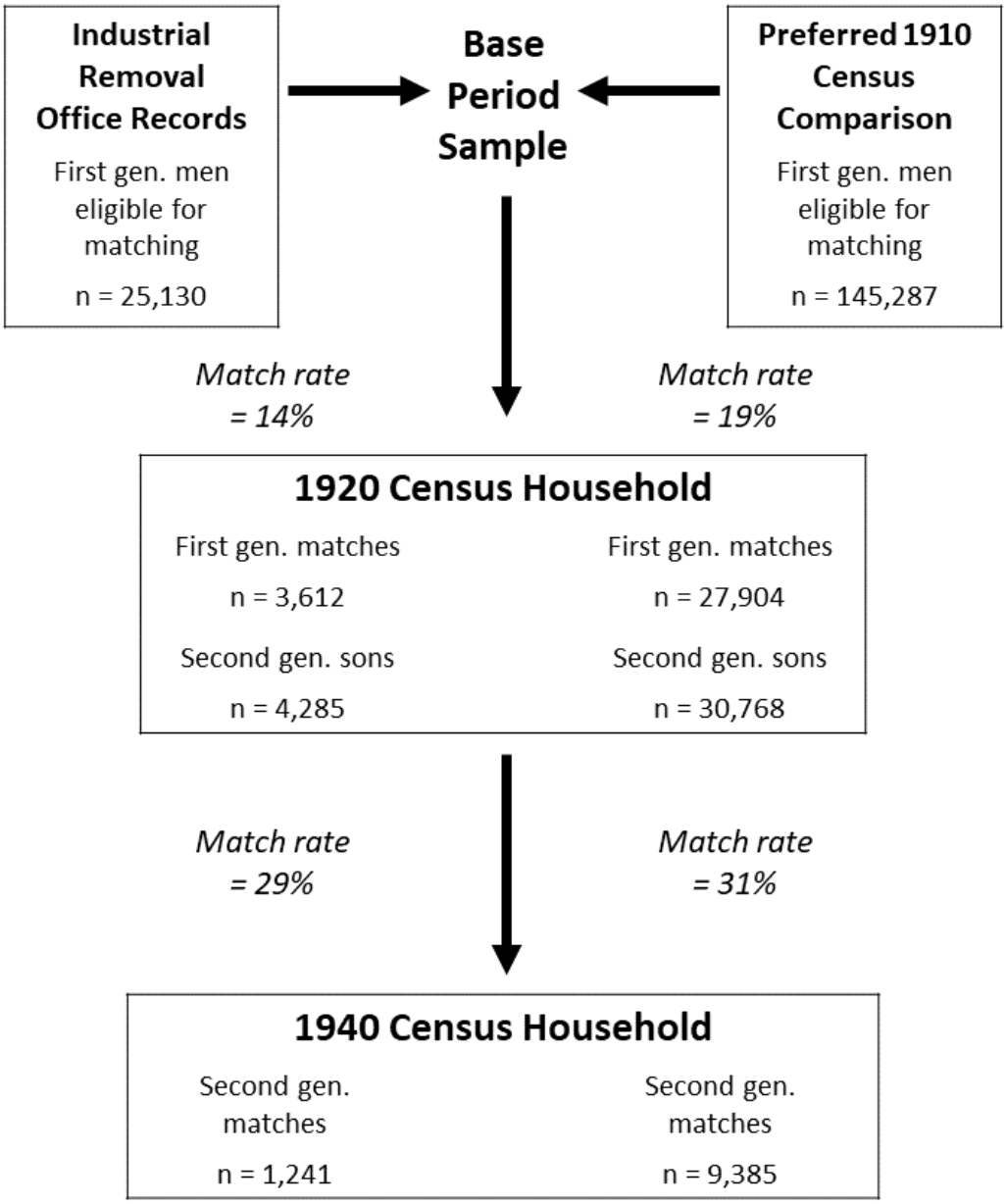
participants. Panel C maps the locations of non-IRO Jewish New Yorkers, our main comparison group, from the 1920 census.

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



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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Data Appendix

DA1. Record linkage & sample construction

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

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

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

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

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

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

We linked our data as follows:

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

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

Data Appendix Table 1. Sample attrition from primary data sources

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

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

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

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

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

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

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

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

DA2. Weighting

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

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

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

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

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

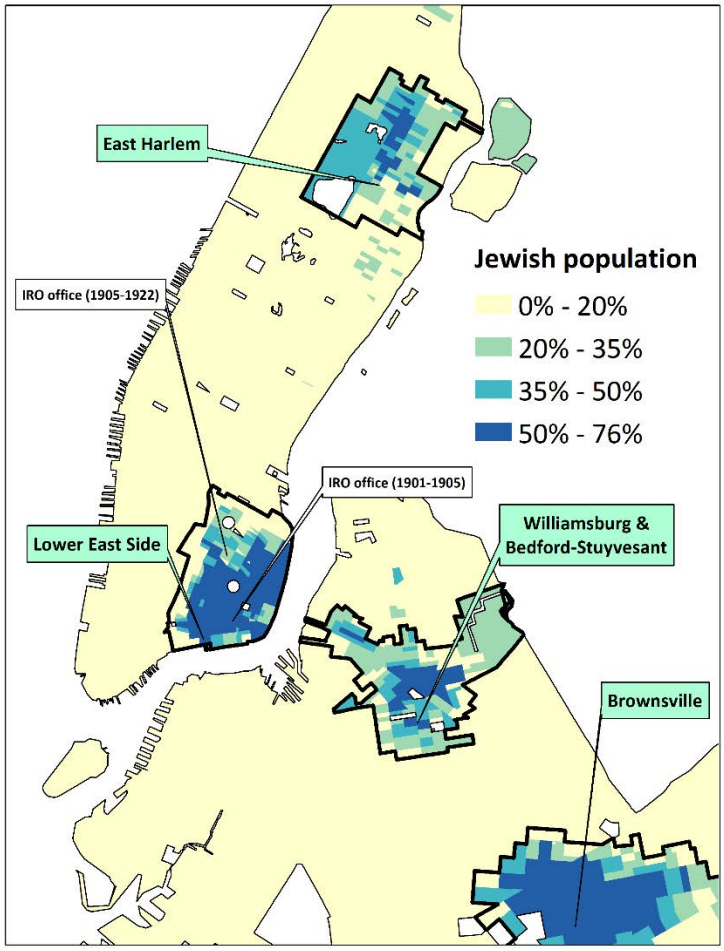
Notes: Sample are foreign-born with a Jewish index > 1.4. Each coefficient is derived from a univariate regression with one independent variable. We suppress the coefficients for small birthplaces categories (e.g. Mexico, Canada, Ireland, England).

DA3. Address matching procedure

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

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

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



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

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

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

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

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

Data Appendix Table 6. Example streets for address matching procedure

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

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

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

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