RF BERLIN

ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 142/26

The Economic Returns to Citizenship: Evidence from a Meta-Analysis

Jens Hainmueller, David Laitin

www.rfberlin.com MAY 2026

The Economic Returns to Citizenship: Evidence from a Meta-Analysis

Authors

Jens Hainmueller, David Laitin

Reference

JEL Codes: C18, F22, J15, J31, J61, K37

Keywords: citizenship; naturalization; immigrant earnings; meta-analysis; selection bias

Recommended Citation: Jens Hainmueller, David Laitin (2026): The Economic Returns to Citizenship: Evidence from a Meta-Analysis. RFBerlin Discussion Paper No. 142/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

The Economic Returns to Citizenship: Evidence from a Meta-Analysis

Jens Hainmueller[a,b] David D. Laitin[a,b]

[a]Department of Political Science, Stanford University
[b]Immigration Policy Lab, Stanford University

May 2026

Abstract. Citizenship is widely summarized as producing a "naturalization premium" for immigrants, but naturalization is highly selected on language, residence, employment, and motivation. We provide a quantitative meta-analysis of citizenship effects on immigrant earnings, synthesizing 63 preferred estimates from 31 studies. Identification strategy explains more between-estimate heterogeneity than citizenship regime, destination country, effect measurement window, outcome data source, earnings-sample basis, or treatment estimand. A multilevel random-effects model yields a pooled earnings premium of 9.6% (95% CI 5.0 to 14.5), broadly consistent with the study-country-collapsed estimate of 11.2% (95% CI 6.1 to 16.5). This average masks a sharp identification-strategy gradient: a randomized encouragement design implies a precisely estimated near-zero effect of 0.4% (95% CI -0.9 to 1.7); the regression-discontinuity study yields 11.3% (95% CI 5.2 to 17.8); reform-based and panel within-person estimates are positive on average but statistically indistinguishable from zero; observational instrumental-variable estimates average 39.9% (95% CI 12.6 to 73.9). Excluding observational instrumental-variable estimates reduces the pooled premium to 5.8% (95% CI 3.6 to 8.1). A within-study diagnostic from the randomized trial shows the same pattern: experimental estimates are near zero, while an observational panel comparison that ignores randomization and uses self-selected citizenship uptake implies a 9.6% gain (95% CI 3.6 to 16.0) four years after the lottery. The published citizenship premium is positive, but its magnitude and causal interpretation are highly sensitive to research design.

JEL Classification: C18, F22, J15, J31, J61, K37.

Keywords: citizenship; naturalization; immigrant earnings; meta-analysis; selection bias.

Acknowledgments: We thank Duncan Lawrence, Dominik Hangartner, and Michael Hotard for useful comments.

Naturalization changes both the rights immigrants hold and the signals they send in the labor market. By granting formal membership, citizenship can remove barriers to work and mobility, stabilize legal status, and alter employers' beliefs about permanence and host-country skills. Together, these channels motivate the expectation of an earnings premium, an expectation that has become the dominant summary of the literature. The National Academies' report on immigrant integration concluded that U.S. citizenship "improves employment outcomes, wage growth, and access to better jobs" (National Academies of Sciences, Engineering, and Medicine, 2015, p. 165). OECD and subsequent reviews likewise summarize naturalization as a route to better economic integration (OECD, 2011; Gathmann and Monscheuer, 2020; Gathmann and Garbers, 2023). Empirical estimates of this "naturalization premium" have grown substantially over four decades, spanning cross-sectional comparisons (Chiswick, 1978; Chiswick and Miller, 1992), panel-data fixed effects (Bratsberg et al., 2002; Engdahl, 2014), policy-reform difference-in-differences (Mazzolari, 2009; Gathmann and Keller, 2018), regression discontinuities (Hainmueller et al., 2019), instrumental variables (Pendakur and Bevelander, 2014), and randomized encouragement (Hainmueller et al., 2026). The literature reports a generally positive premium, but the magnitudes vary widely.

This dispersion in reported premiums matters because naturalization is endogenous to the same factors that determine earnings. Eligibility rules select on residence duration, language, criminal record, and income; application selects on motivation, information, and administrative capacity; approval selects on employment, stability, and institutional networks. Even within-immigrant fixed-effects designs are vulnerable: pre-naturalization labor-market trajectories often slope upward in anticipation of citizenship (Engdahl, 2014; Peters et al., 2018, 2020). As a result, the same population can produce very different estimated premiums depending on what comparison is drawn. Whether the published literature documents a causal effect of citizenship or a selection artifact is therefore a central empirical question for research, theory, and policy.

To answer it, we conduct a quantitative synthesis. We began with a published 53-study appendix bibliography of prior studies on citizenship and immigrant outcomes (Hainmueller et al., 2026). We then conducted forward-citation searches and screened candidate studies through title, abstract, and full text, retaining final published versions whenever available. The earnings sample contains 63 preferred estimates from 31 studies, harmonized to a log-equivalent percent-change metric.

The causal question is counterfactual: would the same immigrant earn more with citizenship than without it? Because we cannot observe earnings for the same person both with and without citizenship, studies must replace the missing counterfactual with an observed comparison. We refer to the source of that comparison, and the variation used to separate citizenship from selection, as the study's identification strategy. This choice matters because strategies differ in how well they address unobserved confounders that may explain earnings differences between naturalized and non-naturalized immigrants. For example, more motivated or better informed immigrants may be both more likely to naturalize and more likely to experience earnings growth even without citizenship. Identification strategy is therefore our central moderator. We distinguish six sources of identifying variation: randomized encouragement designs compare immigrants randomly encouraged or not encouraged to naturalize; regression-discontinuity designs compare applicants on either side of an institutional cutoff; policy or cohort reforms compare immigrants differentially exposed to changes in citizenship access; panel designs compare the same immigrants before and after naturalization; observational instrumental-variable designs use predicted citizenship status from instruments such as eligibility timing or local approval leniency; and cross-sectional designs compare naturalized and non-naturalized immigrants observed at the same time. We benchmark this focal moderator against the other extracted study features that could also shape reported premiums: treatment estimand, effect measurement window, outcome data source, earnings-sample basis, destination country, and citizenship regime measured by the 2019 MIPEX Access to Nationality score.

Pooling across the corpus we recover a positive average premium of approximately ten percent, but the magnitude depends sharply on identification strategy: the randomized study is near zero, the regression-discontinuity study is positive, reform and panel estimates are positive on average but statistically indistinguishable from zero, and observational instrumental-variable estimates are much larger and unstable. The conventional citizenship premium is therefore not a stable causal parameter; it is a design-dependent summary of a highly selected process.

Results

Pooled estimates

Across the 63 preferred earnings estimates from 31 studies (Table S1), a multilevel random-effects model with dataset, study, and within-study sample random effects and CR2-clustered standard errors yields a pooled earnings premium of 9.6% (95% CI 5.0 to 14.5), $P < 0.001$. A two-level model with study and within-study sample random effects gives 9.6% (95% CI 4.7 to 14.7). A study-country-collapsed model that pools to one mean effect per study-country cell yields 11.2% (95% CI 6.1 to 16.5). The three specifications agree: the published literature reports a positive average citizenship premium of approximately one-tenth of earnings. Between-estimate heterogeneity is high (total $\tau^2 = 0.0150$; $I^2 = 99.5\%$), meaning that estimates differ far more than expected from sampling error alone. Most residual dispersion is between studies ($\tau^2 = 0.0108$), with additional clustering among estimates drawn from the same within-study sample ($\tau^2 = 0.0042$). The dataset-family component is near zero ($\tau^2 = 0.0000$), indicating that studies using the same named data source do not move together much once study-level differences are modeled.

Identification-strategy gradient

The pooled average citizenship effect masks a sharp credibility gradient (Fig. 1, Table S2). The single available randomized encouragement design (Hainmueller et al., 2026) implies a precisely estimated near-zero earnings effect of 0.4% (95% CI -0.9 to 1.7). A regression-discontinuity design exploiting Swiss municipal-referendum vote-share thresholds (Hainmueller et al., 2019) yields 11.3% (95% CI 5.2 to 17.8). Policy and cohort-reform estimates pool to 4.7% (95% CI -2.8 to 12.7). Panel within-person estimates pool to 4.2% (95% CI -2.6 to 11.4). Cross-sectional observational estimates pool to 7.6% (95% CI 5.0 to 10.3). Observational instrumental-variables estimates pool to 39.9% (95% CI 12.6 to 73.9), with the wide interval reflecting weak-instrument and exclusion-restriction concerns. Aggregating credibility tiers, the most-credible block (RCT + RDD + policy/cohort reform; 13 estimates from 7 studies) pools to 4.7% (95% CI -0.5 to 10.3); panel within-person estimates (14 estimates from 7 studies), which face known anticipation-effect concerns, pool to 4.2% (95% CI -2.6 to 11.4); and the observational block (cross-sectional + observational instrumental-variable; 36 estimates from 17 studies) pools to 15.8% (95% CI 6.9 to 25.3).

The pattern is consistent with positive selection into naturalization that observational designs may absorb in the citizenship coefficient. The study-collapsed cumulative diagnostic in Fig. 1B shows the same gradient from a second angle. The regression-discontinuity study cautions against a mechanical ranking of designs by effect size, but the broader pattern is clear: as estimates move from cross-sectional status comparisons and observational instruments toward designs closer to the missing counterfactual, the premium shrinks, and in the randomized encouragement study it is essentially zero. The pooled effect remains well below the final study-collapsed cumulative estimate until the observational instrumental-variable tier enters the corpus.

Study-country distribution

The study-country-collapsed forest plot (Fig. 2) confirms that the pooled positive average is not driven by a single paper or a single country cell. Effects span −3% to +71%, with 31 of 35 study-country cells estimating positive effects and 22 estimating statistically significant positive effects at the 5% level. The 35 cells are generated by pooling duplicate preferred earnings rows within each study-destination-country pair; the count exceeds the 31-study count because four multi-country studies contribute two destination-country or regional cells. The largest estimates come disproportionately from observational instrumental-variable and weak observational specifications. The lower end of the distribution includes the randomized encouragement study and several panel within-person studies, while the regression-discontinuity evidence is positive but comes from one institutional setting.

Alternative explanations

The identification-strategy gradient could reflect other study features rather than the source of identifying variation. Figure 3 tests this by fitting competing multilevel meta-regressions and comparing how much estimated between-effect variance each moderator removes relative to the null model (SI Appendix, Table S5). Identification strategy is the dominant one-at-a-time moderator, reducing estimated heterogeneity by $R^2_\tau = 0.690$. The next largest one-at-a-time moderators are effect measurement window ($R^2_\tau = 0.129$) and destination country ($R^2_\tau = 0.107$). Citizenship regime, treatment estimand, outcome data source, and earnings-sample basis explain little additional heterogeneity on their own. The first adjusted bar asks the stricter question: after conditioning on citizenship regime, effect measurement window, outcome data source, earnings-sample basis, and treatment estimand, does identification strategy still matter? It does: adding identification strategy removes a partial $R^2_\tau = 0.739$ of the remaining between-estimate variance. The country-FE adjusted bar repeats the exercise after replacing the country-level citizenship-regime score with pooled destination-country fixed effects; despite the heavier parameter count, identification strategy still removes a partial $R^2_\tau = 0.812$ of residual heterogeneity. Among the moderators we code, identification strategy is therefore by far the most important explanation for variation in reported earnings effects, not simply a proxy for country context, citizenship-regime permissiveness, measurement window, outcome data source, earnings-sample basis, or treatment estimand.

Within-study selection check

A direct within-study design-gradient check comes from Hainmueller et al. (Hainmueller et al., 2026), where the same randomized sample can be analyzed two ways. The experimental contrast uses randomized fee-voucher assignment: immigrants in a lottery were randomly offered a voucher covering the naturalization application fee, which sharply increased naturalization. The observational contrast ignores the randomized assignment and uses citizenship uptake among experimental control-group immigrants who were not encouraged by a voucher, while controlling for common time shocks as in the panel designs common in the literature. Figure 4 applies the Callaway-Sant'Anna difference-in-differences estimator for staggered treatment timing (Callaway and Sant'Anna, 2021) to both contrasts, using never-treated controls and aggregating group-time average treatment effects by event time. The fee-voucher randomization path remains near zero, while the selected post-lottery citizenship path rises to about 9.6% four years after the lottery (95% CI 3.6 to 16.0). A researcher using the randomized encouragement design would therefore conclude that earnings effects are essentially zero, whereas a researcher using selected citizenship uptake in the same panel would conclude that naturalization raises earnings meaningfully. The pre-period estimates in the observational panel are close to zero, so a conventional pretrend diagnostic would not by itself reveal the post-lottery selection bias in this example. This divergence suggests that time-varying selection into naturalization can explain a large part of the gains recovered by observational panel designs. The SI Appendix reports corresponding two-way fixed-effects and fixed-effects counterfactual-imputation versions of the same diagnostic.

Publication bias and significance patterns

Funnel-plot and z-statistic diagnostics (Fig. 5; SI Appendix, Table S4) show no strong corpus-wide publication-bias signature. PET- and PEESE-corrected pooled effects of 8.1% (95% CI 1.9 to 14.6) and 10.4% (95% CI 5.2 to 15.9) are similar in magnitude to the primary 9.6% pool, so the headline premium is not an artifact of publication bias. Within the observational instrumental-variable cell, the largest estimates lie in the upper-right region of the funnel and the z-statistic distribution skews positive; in randomization, regression-discontinuity, and panel cells, neither pattern is visible.

Discussion

This meta-analysis changes the interpretation of the naturalization premium. The positive average in the published literature is a real descriptive regularity, but it should not be treated as a design-invariant causal quantity. Across studies, and within the randomized trial itself, the estimated premium changes when the identifying comparison changes. This pattern points to selection into naturalization as a central part of the empirical object: observational estimates partly describe who naturalizes and when, not only what citizenship does after naturalization.

This conclusion does not imply that citizenship has no economic value. Legal security, labor-market access, and signaling remain credible mechanisms, and some identification-credible estimates are positive. The evidence instead narrows what can be inferred from the conventional premium. It should be read as a design-bound summary of heterogeneous estimates, not as a portable forecast of the causal return to citizenship.

Three implications follow. For empirical research, identification strategy should be a first-order feature of how naturalization-premium estimates are summarized and cited. For policy, forecasts that citizenship reform will generate large earnings gains should be calibrated against identification-credible estimates rather than the observational average alone. For the meta-analytic agenda, the binding constraint is not another correction to the same heterogeneous corpus; it is more plausibly unconfounded variation in citizenship access and uptake. Additional randomized encouragement designs, regression-discontinuity studies, and panel designs with credible counterfactuals would do the most to sharpen the field's answer.

The analysis has limitations. The earnings sample is modest and unevenly distributed across identification strategies: the randomization and regression-discontinuity cells contain one study each. These design cells are therefore also single-context cells, so their pooled estimates should be read as evidence from specific study settings rather than as fully general design averages.

Reflecting where the earnings literature currently exists, the included studies come from Western, high-income destination countries; the resulting estimates therefore say less about citizenship effects in lower-income, non-Western, or otherwise institutionally different contexts. Some estimates require transformation from level or percent effects, with the attendant assumption of an applicable baseline.

Materials and Methods

Search and study selection

Starting from the published appendix list of naturalization-premium studies in Hainmueller et al. (Hainmueller et al., 2026), we conducted forward-citation searches in OpenAlex, Semantic Scholar, and OpenCitations from 1975 onward. The seed list covered citizenship effects on immigrant outcomes broadly, not only earnings. Candidate records were screened on title and abstract using a structured controlled-vocabulary protocol, then on full text. We retained only studies estimating the effect or association of citizenship, naturalization, or access to citizenship on immigrant outcomes; we excluded reverse-direction studies of determinants of naturalization. For this earnings-focused analysis we further required an earnings, wage, salary, or income outcome and a reported or reconstructable uncertainty measure. Final published versions were used wherever available; working papers were retained only when no published version existed. The full search and screening protocol is in *SI Appendix*.

Extraction and coding

Each study contributes one or more estimate-level rows for each substantively distinct outcome × subgroup × treatment estimand × specification combination. Each row records the raw effect, its sign convention, transformations applied to obtain a log-equivalent percent change, and the original-scale uncertainty. When a study reports multiple specifications for the same earnings contrast, the preferred pool uses the authors' primary or preferred estimate; if no preferred estimate is named, we use the most complete main-text specification for the study's central design and broadest eligible sample. The reported moderator analyses use identification strategy plus six additional features: treatment estimand, effect measurement window, outcome data source, earnings-sample basis, destination country, and citizenship regime measured by the 2019 MIPEX Access to Nationality score.

Effect coding

The primary metric is the log-equivalent percent earnings effect: for log earnings or log income coefficients, the reported coefficient is taken as the log effect and reported as $100[\exp(\beta) - 1]$; base-10 log coefficients are converted to natural-log equivalents; level coefficients are converted only when the paper reports a percent effect or when a defensible baseline is available, with standard errors propagated by the delta method. Sign conventions are normalized so that positive values indicate higher earnings under citizenship or citizenship access.

Meta-analytic models and diagnostics

The primary specification is a multilevel random-effects model with random intercepts for dataset family, study, and within-study sample. Eight dataset families pool studies that draw on the same underlying register or survey lineage—Statistics Norway linked registers, Statistics Sweden LINDA/register, Statistics Canada decennial Census, INSEE administrative and tax records, Swiss Labour Force Survey, U.S. decennial Census, American Community Survey, and Current Population Survey—and the remaining studies enter as singleton families (SI Appendix, Section S3). The model is fit by REML with cluster-robust CR2 inference clustered by dataset family. Two-level, study-country-collapsed, and study-collapsed specifications serve as robustness checks; the cumulative design-tier diagnostic pools study-collapsed estimates as increasingly observational designs enter the corpus.

Heterogeneity is examined by stratifying the primary model on identification strategy and by fitting moderator meta-regressions for the main coded study features: identification strategy, citizenship regime, destination country, effect measurement window, outcome data source, earnings-sample basis, and treatment estimand. We summarize moderator models by their proportional reduction in between-estimate variance relative to the null model, and report conditional specifications that add identification strategy after the non-identification moderators, with pooled destination-country fixed effects replacing the country-level MIPEX score in the country-adjusted specification. The within-study Hainmueller et al. diagnostic compares randomized voucher assignment with citizenship uptake among experimental control-group immigrants who were not encouraged, using the Callaway-Sant'Anna difference-in-differences estimator in the balanced credit-bureau panel; the SI Appendix reports two-way fixed-effects and fixed-effects counterfactual-imputation versions. Publication-bias diagnostics include funnel plots, FAT-PET and PEESE corrections (Stanley and Doucouliagos, 2014), and stratified z-statistic distributions following Brodeur, Cook, and Heyes (Brodeur et al., 2020). Full definitions and model output are in the SI Appendix.

Artificial intelligence

The authors used OpenAI Codex, based on GPT-5, and Anthropic Claude Code, based on Claude Opus 4.7, to assist with manuscript preparation, including text editing, consistency checks across the manuscript and SI Appendix, and updates to analysis code, tables, figures, and LaTeX files. The authors reviewed, revised, and fact-checked all AI-assisted output and are solely responsible for the final content. No AI-generated images or graphics are included.

Data Availability

Replication data and code for the analyses reported in this article will be made publicly available.

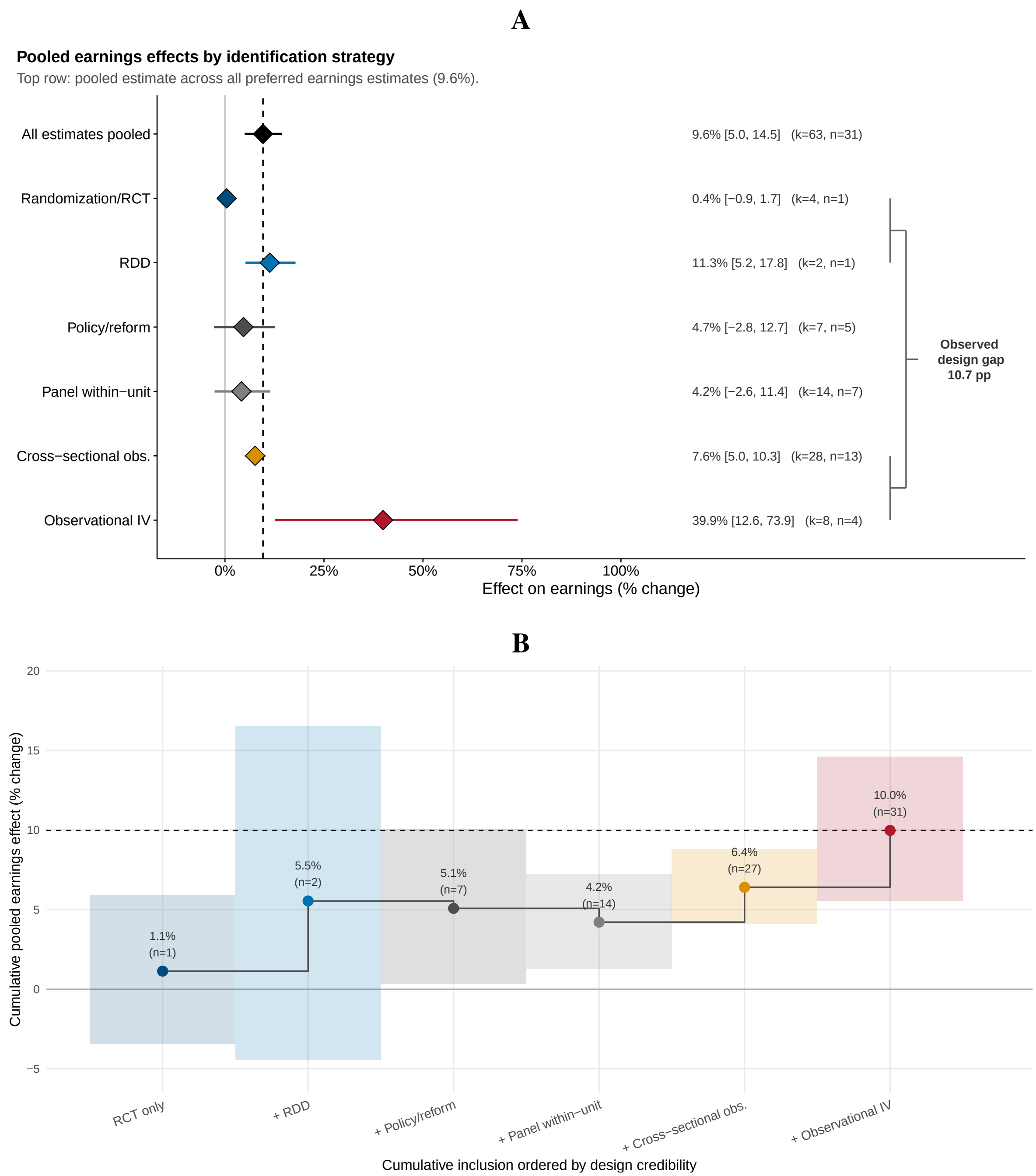

Figure 1. Pooled earnings effects by identification strategy and cumulative design tier. (**A**) The top row shows the pooled estimate across all preferred earnings estimates; remaining rows show pooled premiums by identification strategy. Diamonds show meta-analytic point estimates with 95% CI. The vertical line marks the overall pooled estimate (9.6%). The bracket compares observational identification strategies (cross-sectional + observational instrumental-variable) with strategies that more directly target exogenous variation (randomization + regression discontinuity). In Panel A, k = preferred estimates and n = studies in the row-level multilevel models. (**B**) Study-collapsed cumulative diagnostic. Each study is first collapsed to one effect, then study-level effects are pooled as design tiers are added from more to less credibility-dependent sources of variation. Panel B labels n as the cumulative number of study-level effects, so the count changes as design tiers enter the corpus.

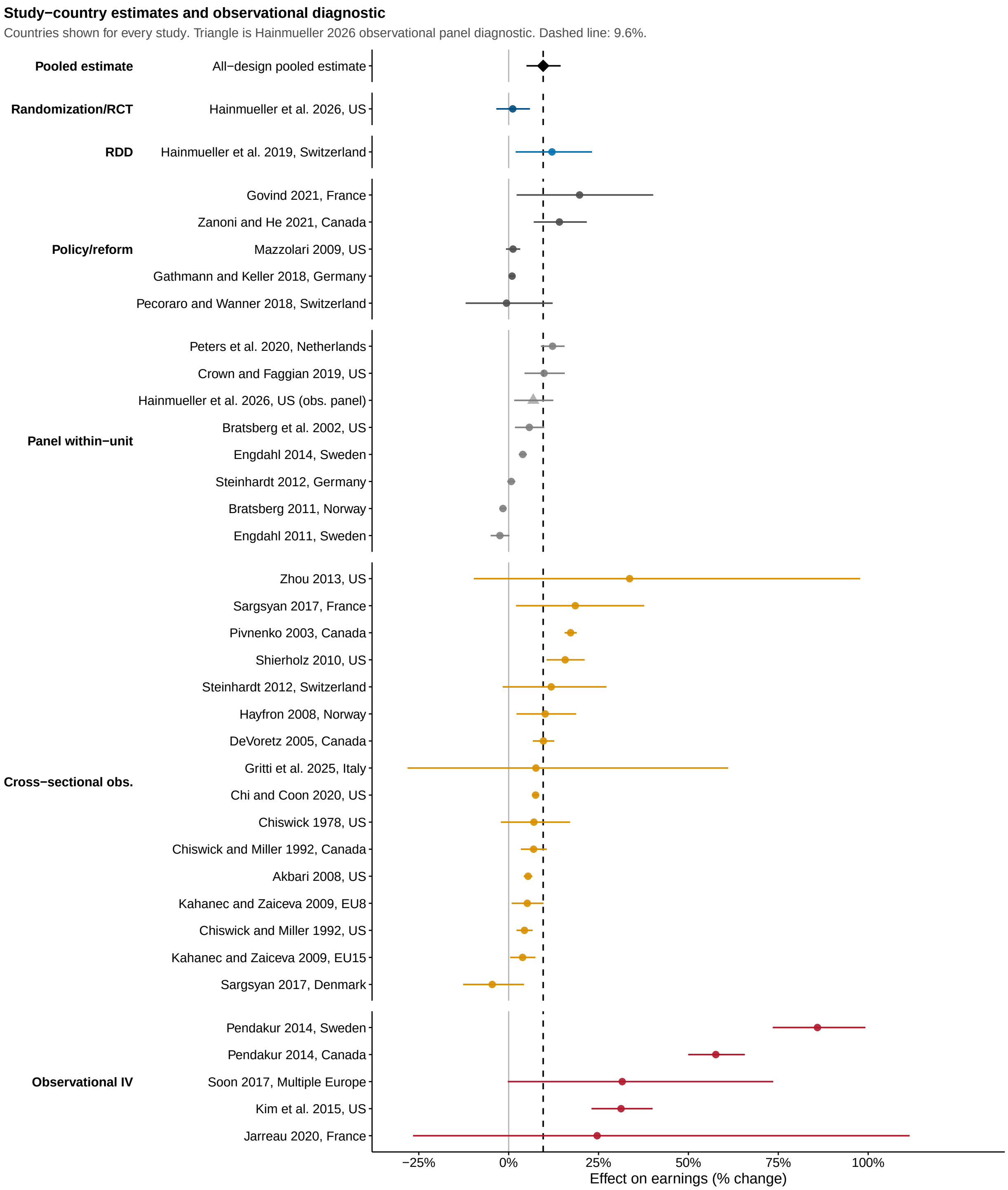

Figure 2. Study-country-collapsed preferred earnings estimates and observational diagnostic. The filled diamond at the top shows the overall pooled estimate across all preferred earnings estimates. Filled circles show one conservative preferred effect per study-country cell, with countries shown in row labels; horizontal bars are 95% CIs. The filled triangle shows the Hainmueller et al. 2026 observational panel diagnostic from the same trial sample. The dashed vertical line marks the overall pooled estimate (9.6%).

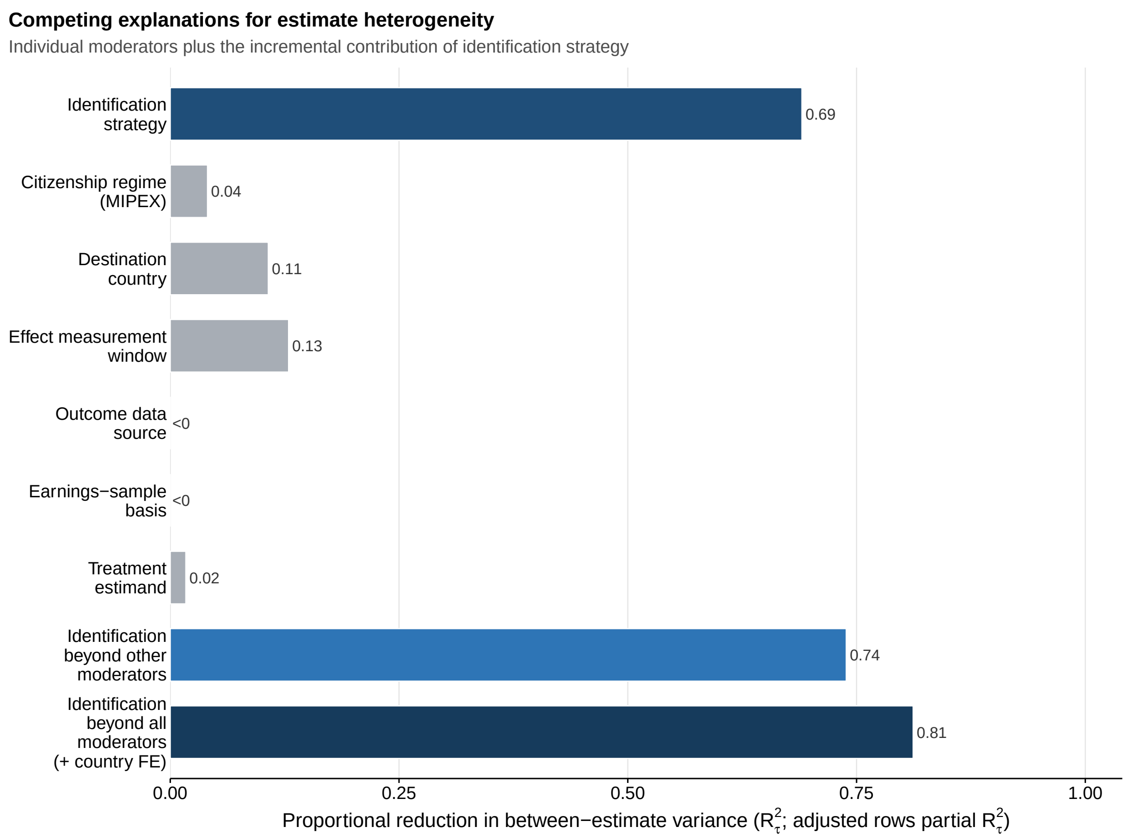

Figure 3. Competing explanations for heterogeneity in reported earnings premiums. Bars show the proportional reduction in between-estimate variance, R^2_τ, from multilevel meta-regressions. The figure reports each moderator separately—identification strategy, citizenship regime (MIPEX Access to Nationality), destination country, effect measurement window, outcome data source, earnings-sample basis, and treatment estimand—then reports the partial R^2_τ from adding identification strategy after the non-identification moderators, first with MIPEX as the citizenship-regime measure and then with pooled destination-country fixed effects replacing MIPEX. Negative estimated reductions are plotted at zero and labeled < 0; exact values appear in SI Appendix, Table S5. Identification strategy explains more heterogeneity than any other single moderator and adds explanatory power beyond the non-identification controls. Computed on the MIPEX-matched subset (57 estimates, 29 studies); full-sample sensitivity reported in SI Appendix.

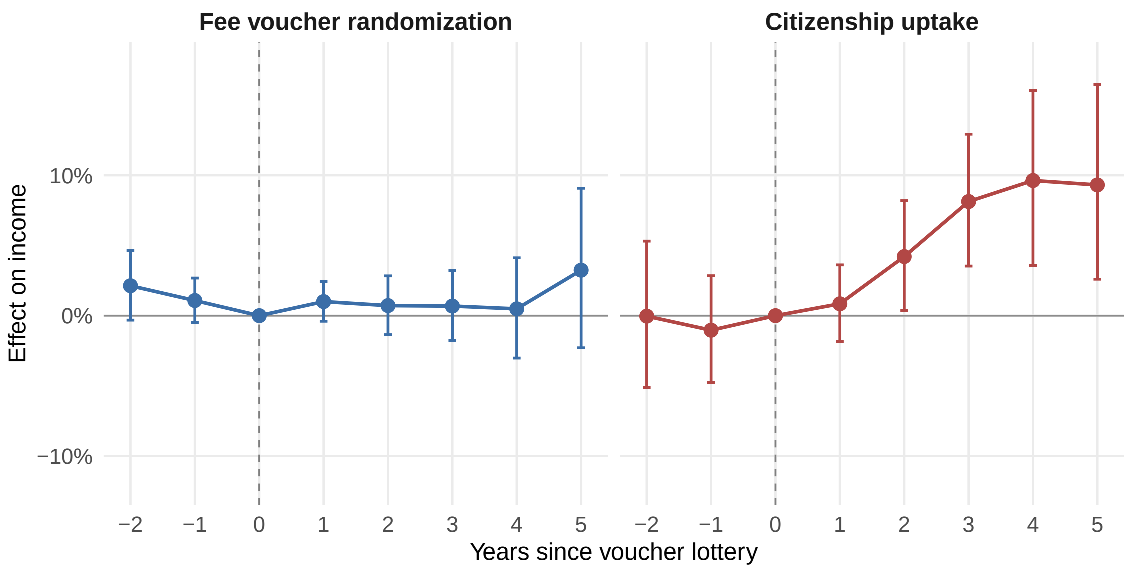

Figure 4. Within-study selection check in Hainmueller et al. (2026). Both panels use the matched six-year credit-bureau panel and the Callaway-Sant'Anna difference-in-differences estimator for staggered treatment timing (Callaway and Sant'Anna, 2021). Year 0 is the lottery year; the income measurement at year 0 is observed immediately before assignment. The left panel estimates randomized voucher-assignment effects. The right panel estimates effects of citizenship uptake among experimental control-group immigrants who were not encouraged. Points show percent effects on income with 95% confidence intervals, transformed as $100\{\exp(\hat{\beta}) - 1\}$. The right panel is a selection diagnostic, not a randomized estimate.

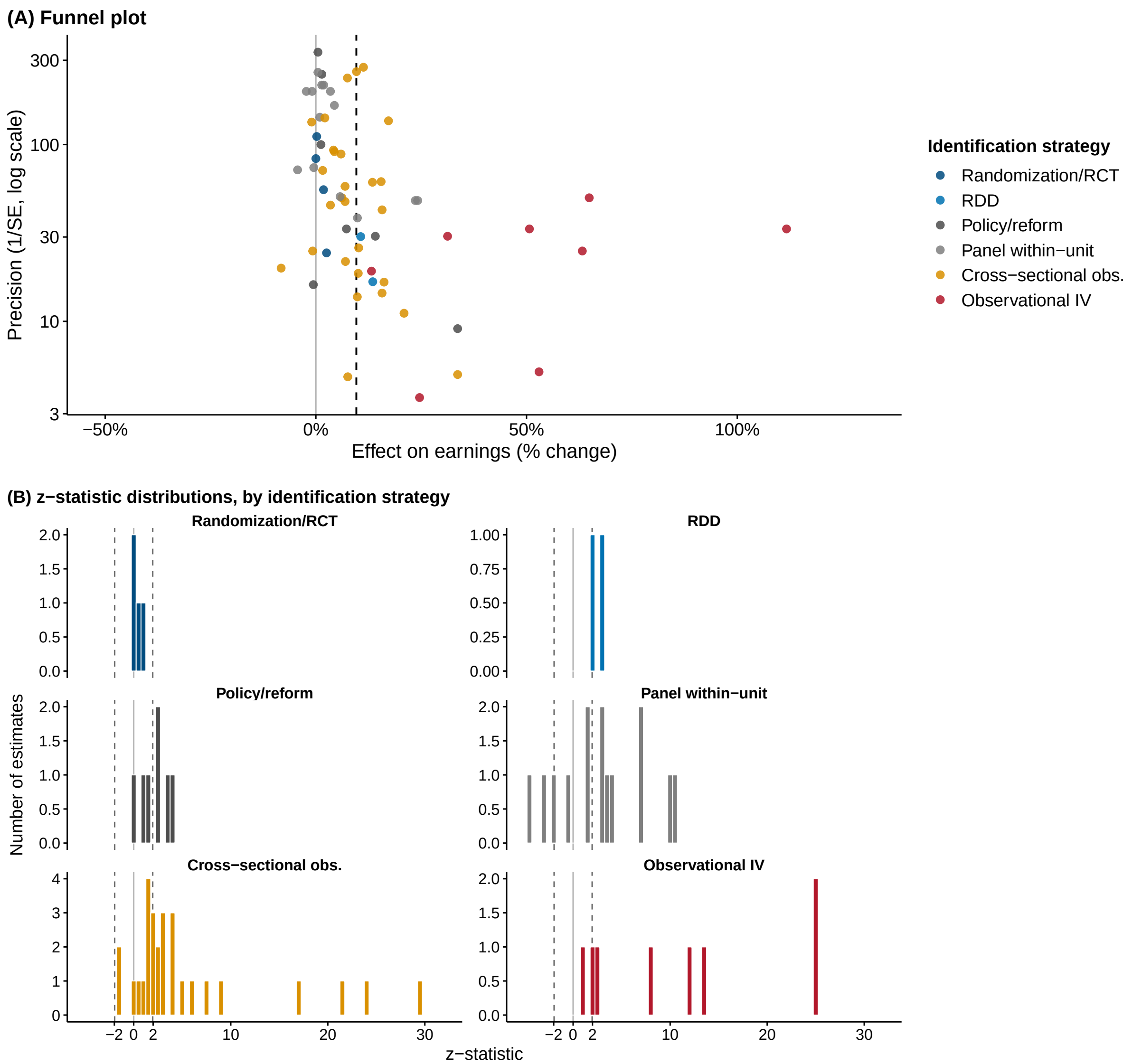

Figure 5. Publication bias and significance patterns by identification strategy. (*A*) Funnel plot, points colored by identification strategy. The vertical line marks the pooled estimate; the diagonal lines are the FAT-PET regression and the symmetric reference. (*B*) z-statistic distributions stratified by identification strategy, with vertical reference at $z = \pm 1.96$. The observational instrumental-variable cell skews positive; randomization, regression-discontinuity, and panel cells do not.

Supporting Information

S1 Search and study selection

Seed corpus

The starting point for the literature search was the appendix list of studies analyzing the naturalization premium published as Table S25 in the supporting information of Hainmueller et al. (2026). That list was constructed by Hainmueller et al. (2026) as a cited-paper and snowball bibliography of the prior citizenship-premium literature: they began with highly cited Google Scholar papers that cited the seminal studies by Chiswick (1978) or Bratsberg et al. (2002) and whose abstracts advertised a statistical test of the returns to citizenship, then expanded the list through references and keyword-based snowballing around immigration and citizenship. The seed list was not restricted to earnings: it included studies of citizenship effects on labor-market, education, health, social-integration, demographic, and other immigrant outcomes. These 53 seed studies span 1978–2024 and cover the major empirical traditions in the citizenship-effects literature: cross-sectional naturalization-status comparisons, panel fixed-effects analyses, policy-reform difference-in-differences, regression discontinuity designs, and observational instrumental variables. The randomized fee-voucher encouragement design in Hainmueller et al. (2026) entered the corpus through the forward-citation and full-text update rather than through the seed list itself.

We treated this seed list as the entry point but did not accept it wholesale. Eleven seed studies were excluded during full-text review: five as review or synthesis articles rather than primary effect-estimation studies (Liebig 2011; Sumption 2012; National Academies 2015; Gathmann 2020; Gathmann 2023); three for lacking any numeric standard errors (Bevelander 2006; Scott 2008; Enchautegui 2015); two for treating naturalization as an outcome or providing descriptive profiles rather than estimating citizenship effects (Mata 1999; Constant 2009); and one duplicate or summary chapter (Steinhardt 2011). Forty-two seed studies survive into the active corpus alongside the forward-citation expansion.

Forward citation search

For each seed study we conducted forward-citation searches in OpenAlex, Semantic Scholar, and OpenCitations COCI, restricted to publication year 1975 or later. Seed studies were matched to citation databases by DOI when available and otherwise by title, author surname, and publication year. The combined search returned 12,285 citing records, which were deduplicated by DOI when available and otherwise by normalized title, first-author surname, and year, yielding 5,116 unique candidate records.

Title and abstract screening

The 5,116 candidates were screened by title and abstract. Inclusion required mention of at least one citizenship/naturalization term, at least one immigrant-population term, and at least one outcome or methodology term. Studies about determinants of naturalization (e.g., interest in citizenship, intentions to naturalize, naturalization rates, or take-up as the dependent variable) and normative or theoretical analyses without empirical estimation were excluded. The screening produced 488 candidates for full-text review.

Additional screening

The 488 first-pass includes were reviewed against the inclusion criteria. This pass retained 45 records for full-text review, sent 44 to manual review, and excluded 347 records, mostly generic immigrant-outcome studies without a citizenship-as-treatment specification or seed-corpus duplicates. A second check of the unclear or excluded candidates identified 7 false-negative includes plus 8 additional records for manual review.

Full-text review

One author adjudicated the 52 records marked for manual review, with 24 included and 28 excluded. Final published versions were used whenever available; working papers were retained only when no published version existed. When supplementary appendices or online supporting materials were available, they were included with the main article for full-text review and extraction. Three studies were further excluded after full-text reading because no usable estimate could be coded, leaving 83 studies in the eligible corpus (42 surviving seed studies plus 41 forward-citation-derived studies).

Earnings restriction

For this earnings-focused meta-analysis we further required that each included study report at least one earnings, wage, salary, or income outcome with a reported or reconstructable uncertainty measure. Among forward-citation records that reached full-text screening but did not enter the earnings pool, Riphahn and Saif (2019) was excluded because the relevant tables report point estimates without standard errors, confidence intervals, or other information sufficient to reconstruct uncertainty. Studies measuring only employment, education, health, social integration, fertility, or other non-earnings outcomes were excluded from the pooled earnings analysis. The earnings sample contains 63 preferred estimates from 31 studies. The full PRISMA-style flow is shown in Figure S1.

S2 Extraction and coding

Coding procedure

For each eligible study, we coded one study-level record and one or more estimate-level records. Estimate rows were defined by substantively distinct combinations of outcome, subgroup, country, treatment estimand, and specification. Each row records the original point estimate and uncertainty measure, the reported model or table source, the treatment contrast, the outcome definition, and the transformations needed to place earnings effects on a common log-equivalent percent scale. When a study reported multiple specifications for the same earnings contrast, the preferred pool used the authors' primary or preferred estimate; if no preferred estimate was named, we used the most complete main-text specification for the study's central design and broadest eligible sample. Coding checks verified required fields, controlled vocabularies, sign conventions, and consistency between point estimates and reported uncertainty.

Coded fields and extracted moderators

For each study and estimate, we coded the features below as moderators in the heterogeneity analyses. The focal moderator is identification strategy: a single categorical variable recording the source of identifying variation.

Extracted moderators.

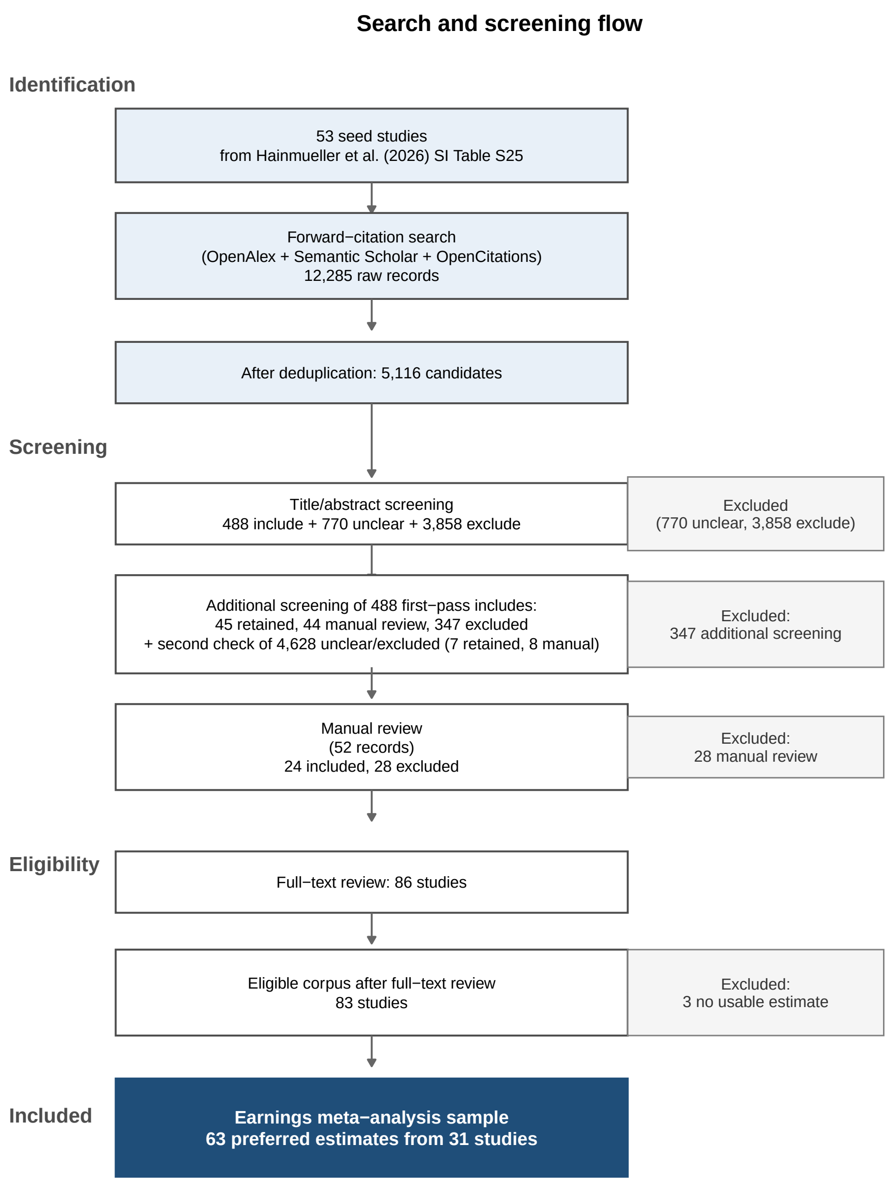

Figure S1. Search and screening flow. The diagram traces the corpus from the 53 seed studies through forward-citation expansion, deduplication, title and abstract screening, full-text review, and the earnings restriction.

Moderator	Coding
Identification strategy	Randomization, regression discontinuity, policy or cohort reform, panel within-person, observational instrumental variables, or cross-sectional observational.
Treatment estimand	Actual citizenship/naturalization versus access to citizenship through eligibility, policy exposure, or randomized encouragement.
Effect measurement window	Time between the relevant treatment or exposure and the earnings outcome. For the moderator analyses, we code this as a categorical factor with four main categories: contemporaneous, short-term follow-up (> 0–2 years), medium-term follow-up (> 2–5 years), and long-term follow-up (> 5 years). Cross-sectional citizenship-status comparisons enter the contemporaneous category when citizenship status and earnings are measured at the same time; short-term is reserved for post-treatment outcomes.
Outcome data source	Two-category indicator for administrative versus self-reported earnings data. Administrative data include linked tax, payroll, credit, and register records; census and survey earnings are coded as self-reported.
Earnings-sample basis	Two-category indicator for whether the estimate uses a broad income sample that includes zero earners or household income, or restricts to employed workers or positive earners.
Destination country	Destination country for the estimate. Most studies contribute estimates for a single destination country; in multi-country studies, destination country is coded at the estimate level. One pooled cross-country estimate is coded as multi-country.
Citizenship regime	2019 MIPEX Access to Nationality score (Solano and Huddleston, 2020), a country-level index of how inclusive or restrictive access to citizenship is. Higher scores indicate more inclusive access.

Outcome harmonization

The primary effect-size metric is a log-equivalent percent earnings effect computed as $100[\exp(\beta)-1]$ for log-earnings or log-income coefficients. Base-10 log coefficients are converted to natural-log equivalents by dividing by $\log_{10}(e)$. Level coefficients are converted to log-equivalent percent only when the paper reports a percent effect or when a defensible baseline mean is available; otherwise the row is excluded from the pooled earnings analysis. Standard errors for converted percent effects are obtained by the delta method.

Sign convention

All effects are coded so that a positive sign means citizenship, naturalization, or access to citizenship increases earnings relative to the comparison group. When a paper's reported coefficient uses the reverse contrast (e.g., longer waiting period as the regressor), the sign is reversed before applying transformations and the sign flip is documented per row.

S3 Statistical model

Primary specification

The primary pooled estimate uses a multilevel random-effects meta-analysis (`rma.mv` from the R package `metafor`) on log-equivalent percent earnings effects:

$$y_{dsi} = \mu + u_d + u_s + u_{si} + \varepsilon_{dsi},$$

where y_{dsi} is the ith estimate in study s from dataset family d. Dataset families group studies that draw on the same underlying register or survey lineage. The term $u_d \sim N(0, \tau_d^2)$ is a dataset-family random intercept, $u_s \sim N(0, \tau_s^2)$ is a study random intercept, u_{si} captures within-study sample dependence, and ε_{dsi} has variance $\sigma_{dsi}^2 = \mathrm{SE}_{dsi}^2$. Variance components are estimated by REML. Inference uses cluster-robust standard errors of CR2 type clustered at the dataset-family level (`robust` from `clubSandwich`).

We code eight dataset families based on the data lineage reported in each study; remaining studies enter as singleton families.

- **Norway register** (Statistics Norway FD-Trygd / linked administrative registers): Hayfron (2008), Bratsberg, Raaum, and Røed (2011).
- **Sweden register** (Statistics Sweden LINDA / linked register data): Engdahl (2011), Engdahl (2014), and the Swedish arm of Pendakur and Bevelander (2014).
- **Canada Census** (Statistics Canada decennial Census PUMF): the Canadian arm of Chiswick and Miller (1992), Pivnenko and DeVoretz (2003), DeVoretz and Pivnenko (2005), and the Canadian arm of Pendakur and Bevelander (2014).
- **France administrative** (INSEE DADS / EDP / tax records): Jarreau (2020), Govind (2021).
- **Switzerland LFS** (Swiss Labour Force Survey, SAKE): Steinhardt (2012, Switzerland), Pecoraro and Wanner (2018).
- **U.S. Census** (U.S. decennial Census IPUMS microdata): Chiswick (1978), the U.S. arm of Chiswick and Miller (1992), Akbari (2008), Mazzolari (2009).
- **U.S. ACS** (American Community Survey IPUMS microdata): Zhou and Lee (2013), Chi and Coon (2020).
- **U.S. CPS** (Current Population Survey March supplement): Shierholz (2010), Kim et al. (2015).

Other studies—including Steinhardt (2012, Germany; IAB register), Gathmann and Keller (2018; German Microcensus), Sargsyan (2017; Luxembourg Income Study), Zanoni and He (2021; Canadian Labour Force Survey), Bratsberg, Ragan, and Nasir (2002; mixed NLSY/Census/CPS), Crown and Faggian (2019; NSF Survey of Doctoral Recipients), Hainmueller et al. (2019; Swiss municipal-referendum records), Hainmueller et al. (2026; voucher-lottery and credit-bureau records), Peters et al. (2020; Netherlands register), Gritti et al. (2025; Italian linked data), Soon (2017; multi-country survey), and Kahanec and Zaiceva (2009; EU15/EU8 survey aggregates)—enter as singleton families because they draw on data lineages not shared with another study in the corpus.

Estimated variance components for the primary model are $\tau_d^2 = 0.0000$ at the dataset-family level, $\tau_s^2 = 0.0108$ at the study level, and $\tau_{si}^2 = 0.0042$ within studies (total $\tau^2 = 0.0150$). We report $I^2 = 99.5\%$ using the generalized Higgins-Thompson formula $I^2 = \tau^2/(\tau^2 + v_{\text{typical}})$, with $v_{\text{typical}} = (k-1)\sum w_i/[(\sum w_i)^2 - \sum w_i^2]$ and $w_i = 1/\text{vi}$. The dataset-family component is small because after coding eight families explicitly, the residual covariance among studies in the same family is largely absorbed by the study-level random intercept; we retain the level for transparency and for the clustering of CR2 standard errors.

Robustness specifications

Two-level (study + within-study sample), study-collapsed, and study-country-collapsed specifications are reported alongside the primary model. The primary row-level model uses all 63 preferred estimates from 31 studies. The study-collapsed specification uses one inverse-variance summary per study (31 observations). The study-country-collapsed specification uses one inverse-variance summary per study-destination-country pair (35 observations), because Chiswick and Miller (1992), Kahanec and Zaiceva (2009), Pendakur and Bevelander (2014), and Sargsyan (2017) each contribute two destination-country or regional cells.

Stratified pooling

Per-identification-strategy pooled estimates are computed with the same multilevel specification on the corresponding subsample, with sufficient k for variance estimation. Cells with $k < 3$ are reported as point estimates with their reported standard error rather than as pooled meta-analytic estimates.

Alternative-moderator robustness check

To test whether the identification-strategy gradient is confounded by other study features, we fit multilevel meta-regressions using the preferred row-level earnings estimates and the moderator fields defined above. Citizenship regime is measured with the continuous 2019 MIPEX Access to Nationality score from the Migrant Integration Policy Index 2020 (Solano and Huddleston, 2020). Destination country enters as fixed effects, with sparse countries pooled for estimation. These moderator-comparison models are fit by maximum likelihood (ML) rather than REML so that the variance components and likelihood-based information criteria are comparable across specifications with different fixed-effect terms; the primary pooled model in Section S3 uses REML.

Table S5 reports each moderator one at a time and then two conditional models. The first adds identification strategy after citizenship regime, effect measurement window, outcome data source, earnings-sample basis, and treatment estimand. The second repeats this calculation after replacing the country-level MIPEX score with pooled destination-country fixed effects. The country-inclusive specification is parameter-heavy (15 and 20 model parameters before and after adding identification strategy, respectively, on 57 estimates), but it converges and yields a similar conclusion: adding identification strategy reduces residual between-estimate variance by partial $R^2_\tau = 0.812$. We summarize explanatory power as the proportional reduction in total estimated between-estimate variance, R^2_τ, relative to the null multilevel model. For the adjusted rows, partial R^2_τ is the proportional reduction in residual between-estimate variance from adding identification strategy after the corresponding non-identification baseline.

The MIPEX-matched sample contains 57 estimates from 29 studies because two studies report estimates for multi-country aggregates with no 2019 MIPEX score: Kahanec and Zaiceva (2009), which reports separate EU15 and EU8 estimates, and Soon (2017), which pools multiple European host countries. As a sensitivity check, we refit the moderator horse race on the full primary sample of 63 estimates from 31 studies, omitting the MIPEX moderator (whose merge requires a single destination country). Identification strategy still dominates: $R^2_\tau = 0.607$ on its own and a partial $R^2_\tau = 0.659$ after the remaining non-identification moderators (effect measurement window, outcome data source, earnings-sample basis, and treatment estimand). The relative ranking of moderators is unchanged.

Within-study selection diagnostic

For Hainmueller et al. (2026), the randomized voucher estimates are the preferred rows in the meta-analysis. To show how selection can re-enter when the experimental design is ignored, the main text applies the Callaway-Sant'Anna difference-in-differences estimator for staggered treatment timing (Callaway and Sant'Anna, 2021). Figure S2 reports the corresponding two-way fixed-effects event-time specification using the authors' replication data and code. Figure S3 repeats the side-by-side comparison using the fixed-effects counterfactual-imputation estimator of Liu et al. (2024). This estimator fits a two-way fixed-effects outcome model using untreated observations, imputes untreated potential outcomes for treated observations, and averages observed-minus-imputed gaps by event time. The main-text figure and both event-time figures restrict the matched six-year credit-bureau panel to individuals with observed income in all six panel years. Results are substantively similar across the three estimators: the randomized voucher path remains close to zero, while the selected citizenship path rises after treatment. In both SI figures, the left panel estimates randomized voucher effects and the right panel estimates effects of citizenship uptake among experimental control-group immigrants who were not encouraged. Table S6 reports the experimental income estimates alongside observational panel diagnostics from the same credit-bureau panel. The selected-status rows are shown only as the within-study diagnostic.

Publication-bias diagnostics

FAT-PET and PEESE corrections (Stanley and Doucouliagos, 2014) are computed on the study-country-collapsed dataset. The FAT-PET regression is $\text{effect}_s = \beta_0 + \beta_1 \text{SE}_s + \varepsilon_s$, weighted by inverse variance, with the publication-bias-corrected pooled estimate given by β_0. PEESE is $\text{effect}_s = \beta_0 + \beta_1 \text{SE}_s^2 + \varepsilon_s$. Funnel plots and z-statistic distributions stratified by identification strategy follow Brodeur et al. (2020).

S4 Supplementary tables

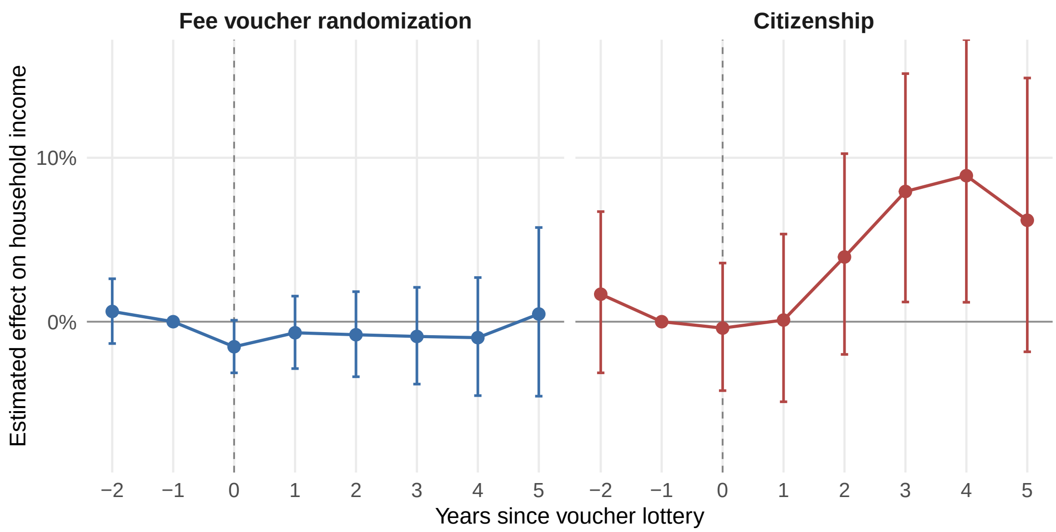

Figure S2. Side-by-side event-time difference-in-differences estimates in Hainmueller et al. (2026). The left panel estimates voucher-assignment effects in the matched six-year credit-bureau panel, restricted to individuals with observed income in all six panel years, using individual and calendar-year fixed effects, the original randomization-block weights, and standard errors clustered by individual. The right panel estimates effects of citizenship uptake among experimental control-group immigrants who were not encouraged. The omitted reference period is one year before the lottery; the vertical dashed line marks the lottery year. Coefficients are transformed to percent effects using $100\{\exp(\hat{\beta}) - 1\}$. The right panel is a selection diagnostic, not a randomized estimate.

Figure S3. Fixed-effects counterfactual-imputation estimates in Hainmueller et al. (2026). Both panels use the matched six-year credit-bureau panel, restricted to individuals with observed income in all six panel years, and the fixed-effects counterfactual-imputation estimator of Liu et al. (2024), implemented in the R package `fect`. The left panel estimates randomized voucher-assignment effects; the right panel estimates effects of citizenship uptake among experimental control-group immigrants who were not encouraged. Coefficients are transformed to percent effects using $100\{\exp(\hat{\beta}) - 1\}$. The pattern is similar to the Callaway-Sant'Anna estimates in the main text: the randomized voucher path remains close to zero, while the selected citizenship path rises after treatment.

Table S1. Studies Contributing Preferred Earnings Estimates

Study	Country	Design	Estimand	Outcome source	Earnings sample	Provenance	k	Effect [95% CI]
Chiswick 1978 (1978)	United States	Cross-sectional obs.	status association	census	positive earners only	Footnote 26, regression 2 (OLS)	1	7.0 [-2.2, 17.1]
Chiswick and Miller 1992 (1992)	Canada; United States	Cross-sectional obs.	status association	census	employed workers only	Table 7-8, column b (OLS); Table 7-9, column b (OLS)	2	5.7 [2.7, 8.7]
Bratsberg et al. 2002 (2002)	United States	Panel within-unit	status association	survey self report	employed workers only	Table 5, column 1 (FE)	1	5.7 [1.7, 9.9]
Pivnenko 2003 (2003)	Canada	Cross-sectional obs.	status association	census	positive earners only	Table 4 (OLS)	1	17.2 [15.6, 18.9]
DeVoretz 2005 (2005)	Canada	Cross-sectional obs.	status association	census	positive earners only	Table 3, column 1 (OLS); Table 3, column 2 (OLS); Table 3, column 3 (OLS); Table 3, column 4 (OLS)	4	9.7 [6.7, 12.7]
Akbari 2008 (2008)	United States	Cross-sectional obs.	status association	census	employed workers only	Table 3, column 1 (OLS); Table 3, column 2 (OLS); Table 3, column 3 (OLS); Table 3, column 4 (OLS)	4	5.4 [4.2, 6.6]
Hayfron 2008 (2008)	Norway	Cross-sectional obs.	status association	administrative register	employed workers only	Table 3, column 6; text calculation below Table 3 (RE)	1	10.2 [2.2, 18.8]
Kahanec and Zaiceva 2009 (2009)	EU15; EU8	Cross-sectional obs.	status association	survey self report	employed workers only	Table IV, column 3 (OLS); Table IV, column 4 (OLS); Table IV, column 7 (OLS); Table IV, column 8 (OLS)	4	4.5 [0.6, 8.6]
Mazzolari 2009 (2009)	United States	Policy/reform	itt access	census	positive earners only	Table 6, Panel A (DiD)	1	1.2 [-0.8, 3.2]
Shierholz 2010 (2010)	United States	Cross-sectional obs.	status association	survey self report	household income all respondents	Table 3, column 3 (OLS)	1	15.7 [10.5, 21.2]
Bratsberg 2011 (2011)	Norway	Panel within-unit	status association	administrative register	employed workers only	Table 7.4, Model III, All (FE)	2	-1.6 [-2.5, -0.6]
Engdahl 2011 (2011)	Sweden	Panel within-unit	status association	administrative register	positive earners only	Table 3.A2.8 (panel); Table 3.A2.9 (panel)	2	-2.4 [-5.0, 0.2]
Steinhardt 2012 (2012)	Switzerland	Cross-sectional obs.	status association	survey self report	employed workers only	Table 2 (decomposition); Table 4 (decomposition)	3	11.8 [-1.7, 27.2]
Steinhardt 2012 (2012)	Germany	Panel within-unit	status association	administrative register	employed workers only	Table 4 (FE)	2	0.7 [-0.4, 1.8]
Zhou 2013 (2013)	United States	Cross-sectional obs.	status association	census	employed workers only	Table 3 (OLS)	1	33.6 [-9.7, 97.8]
Engdahl 2014 (2014)	Sweden	Panel within-unit	status association	administrative register	positive earners only	Table 4, Model III (FE)	2	3.9 [2.8, 5.1]
Pendakur 2014 (2014)	Canada; Sweden	Observational IV	late citizenship	administrative register; census	positive earners only	Table 3 (IV/2SLS)	4	71.2 [61.1, 81.8]
Kim et al. 2015 (2015)	United States	Observational IV	ate citizenship	survey self report	household income all respondents	Table 4, earnings equation (Treatment-effect model using family income...)	1	31.3 [23.0, 40.0]
Sargsyan 2017 (2017)	Denmark; France	Cross-sectional obs.	status association	survey self report	employed workers only	Table 4 (decomposition)	4	6.4 [-6.0, 20.3]
Soon 2017 (2017)	Multiple European host countries	Observational IV	itt access; late citizenship	survey self report	household income all respondents	Table IV, Specification 1 (IV/2SLS); Table V, Specification 1 (IV/2SLS)	2	31.6 [-0.3, 73.6]
Gathmann and Keller 2018 (2018)	Germany	Policy/reform	eligibility or policy effect	census	employed workers only	Table 3, column 8 and Section 4.2 text (Reduced-form eligibility regression, Table...); Table 3, column 4 (Reduced-form eligibility regression, Table...)	2	1.0 [0.3, 1.7]
Pecoraro and Wanner 2018 (2018)	Switzerland	Policy/reform	eligibility or policy effect	survey self report	employed workers only	Table 1 (DiD)	1	-0.6 [-12.0, 12.2]
Crown and Faggian 2019 (2019)	United States	Panel within-unit	ate citizenship	survey self report	employed workers only	Table 4, column 1 (OLS)	1	9.9 [4.4, 15.6]
Hainmueller et al. 2019 (2019)	Switzerland	RDD	itt access	administrative register	includes zero earnings	Supplementary Table S6, column 6; Figure 1D and Results text (RDD); Supplementary Table S5, column 6; Figure 1D and Results text (DiD)	2	12.1 [1.9, 23.2]
Chi and Coon 2020 (2020)	United States	Cross-sectional obs.	status association	census	positive earners only	Table 3, column 1 (OLS)	1	7.5 [6.6, 8.4]
Jarreau 2020 (2020)	France	Observational IV	late citizenship	tax record	employed workers only	Table 9, column 3 (IV/2SLS)	1	24.6 [-26.6, 111.5]
Peters et al. 2020 (2020)	Netherlands	Panel within-unit	status association	administrative register	employed workers only; includes zero earnings	Table 1, Model 1 (FE); Table 1, Model 2 (FE)	4	12.2 [8.9, 15.6]
Govind 2021 (2021)	France	Policy/reform	itt access	administrative register	employed workers only	Table 3 (DiD)	2	19.7 [2.2, 40.2]
Zanoni and He 2021 (2021)	Canada	Policy/reform	itt access	survey self report	employed workers only	Table 5, Column 1, Policy Implementation (DiD)	1	14.1 [7.0, 21.7]
Gritti et al. 2025 (2025)	Italy	Cross-sectional obs.	ate citizenship	linked data	positive earners only	Figure 6, full sample, 2014, + Sociodemo t-1 (OLS)	1	7.6 [-28.1, 61.1]
Hainmueller et al. 2026 (2026)	United States	Randomization/RCT	itt access; late citizenship	credit record	household income all respondents	Table 2, Panel A (ITT); Table 2, Panel B (IV/2SLS); Table 3, Panel A (ITT); Table 3, Panel B; SI Appendix Table S4, Panel B (IV/2SLS)	4	1.1 [-3.4, 5.9]

Notes: The provenance column identifies the table, figure, panel, column, or specification from which the preferred row(s) were extracted. Short model tags are OLS, FE (fixed effects), RE (random effects), DiD, RDD, ITT, IV/2SLS, panel, and decomposition. When $k > 1$, multiple preferred rows from the same study are collapsed into one study-level entry.

Table S2. Pooled effects by identification strategy

Design	k	Studies	Effect [95% CI]	p
Randomization/RCT	4	1	0.4 [-0.9, 1.7]	0.530
RDD	2	1	11.3 [5.2, 17.8]	<0.001
Policy/reform	7	5	4.7 [-2.8, 12.7]	0.155
Panel within-unit	14	7	4.2 [-2.6, 11.4]	0.173
Observational IV	8	4	39.9 [12.6, 73.9]	0.002
Cross-sectional obs.	28	13	7.6 [5.0, 10.3]	<0.001

Table S3. Sign and significance by identification strategy

Design	Positive p<0.05	Positive n.s.	Negative p<0.05	Negative n.s.
Randomization/RCT	0	3	0	0
RDD	2	0	0	0
Policy/reform	4	2	0	1
Panel within-unit	8	2	2	2
Observational IV	7	1	0	0
Cross-sectional obs.	18	7	0	3

Table S4. PET and PEESE diagnostics

Diagnostic	k	Effect [95% CI]	p
PET	35	8.1 [1.9, 14.6]	0.010
PEESE	35	10.4 [5.2, 15.9]	<0.001

Notes: PET and PEESE are study-country-collapsed diagnostics; interpret cautiously because effects are heterogeneous.

Table S5. Moderator contribution to explained heterogeneity

Model	k	Studies	Parameters	R^2_τ	Partial R^2_τ
No moderators	57	29	1	0.000	
Identification strategy	57	29	6	0.690	
Citizenship regime (MIPEX)	57	29	2	0.041	
Destination-country fixed effects	57	29	9	0.107	
Effect measurement window	57	29	4	0.129	
Outcome data source	57	29	2	-0.022	
Earnings-sample basis	57	29	2	-0.013	
Treatment estimand	57	29	2	0.017	
Identification strategy added after other moderators	57	29	13	0.805	0.739
Identification strategy added after other moderators plus country FE	57	29	20	0.897	0.812

Notes: Models use the row-level preferred earnings estimates matched to the 2019 MIPEX Access to Nationality score. R^2_τ is the proportional reduction in total estimated between-estimate variance relative to the null multilevel model. The first seven non-null rows show each reported moderator one at a time: identification strategy, citizenship regime (MIPEX Access to Nationality), destination country, effect measurement window, outcome data source, earnings-sample basis (broad income sample versus employed/positive earners), and treatment estimand. The first adjusted row adds identification strategy after the non-identification moderators: citizenship regime, outcome data source, earnings-sample basis, effect measurement window, and treatment estimand. Its partial R^2_τ is the proportional reduction in residual between-estimate variance relative to that omitted non-identification baseline. The final row repeats this calculation after replacing the country-level MIPEX score with pooled destination-country fixed effects. Destination-country fixed effects pool countries with fewer than three row-level estimates into an other-country category.

Table S6. Within-study diagnostic in Hainmueller et al. (2026)

Identification source	Estimated contrast	Sample and model	Effect [95% CI]	p
Voucher randomization	ITT: voucher access	Three-year credit-bureau income	0.0 [-2.3, 2.4]	1.000
Voucher randomization	LATE: citizenship uptake	Three-year credit-bureau income	2.5 [-5.4, 11.1]	0.542
Voucher randomization	ITT: voucher access	Six-year panel with individual and year fixed effects	0.2 [-1.6, 2.0]	0.824
Voucher randomization	LATE: citizenship uptake	Six-year panel with individual and year fixed effects	1.8 [-1.7, 5.5]	0.317
Selected application timing	Static panel: submitted N-400 vs. not	Control group only; individual and year fixed effects	3.7 [0.1, 7.4]	0.043
Selected application timing	Dynamic panel: effect after 4 years	Control group only; individual and year fixed effects	6.8 [1.5, 12.4]	0.011
Selected application timing	Dynamic panel: effect after 4 years	Full sample; individual and year fixed effects	6.1 [2.3, 10.0]	0.001

Notes: Outcome is log household income from the credit-bureau panel, reported as percent changes using $100\{\exp(\hat{\beta}) - 1\}$. Experimental rows use the voucher lottery or voucher-induced citizenship uptake. Observational rows ignore the randomized assignment and use selected N-400 submission timing; they are coded as selection diagnostics and excluded from the preferred meta-analytic pool.

References

Ather H. Akbari. Immigrant naturalization and its impacts on immigrant labour market performance and treasury. In Pieter Bevelander and Don J. DeVoretz, editors, *The Economics of Citizenship*, pages 129–154. Holger Press, Malmö, 2008.

Bernt Bratsberg and Oddbjørn Raaum. The labour market outcomes of naturalised citizens in Norway. In *Naturalisation: A Passport for the Better Integration of Immigrants?* OECD Publishing, Paris, 2011. doi: 10.1787/9789264099104-10-en.

Bernt Bratsberg, James F. Ragan, and Zafar M. Nasir. The effect of naturalization on wage growth: A panel study of young male immigrants. *Journal of Labor Economics*, 20(3):568–597, 2002. doi: 10.1086/339616.

Abel Brodeur, Nikolai Cook, and Anthony Heyes. Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11):3634–3660, 2020. doi: 10.1257/aer.20190687.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.

Miao Chi and Michael Coon. Variations in naturalization premiums by country of origin. *Eastern Economic Journal*, 46(1):102–125, 2020. doi: 10.1057/s41302-019-00149-0.

Barry R. Chiswick. The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86(5):897–921, 1978. doi: 10.1086/260717.

Barry R. Chiswick and Paul W. Miller. Language in the immigrant labor market. In Barry R. Chiswick, editor, *Immigration, Language and Ethnicity: Canada and the United States*, pages 229–296. American Enterprise Institute, Washington, DC, 1992.

Daniel Crown and Alessandra Faggian. Naturalization and the productivity of foreign-born doctorates. *Journal of Geographical Systems*, 21(4):533–556, 2019. doi: 10.1007/s10109-019-00301-6.

Mattias Engdahl. The impact of naturalisation on labour market outcomes in Sweden. In *Naturalisation: A Passport for the Better Integration of Immigrants?* OECD Publishing, Paris, 2011. doi: 10.1787/9789264099104-6-en.

Mattias Engdahl. Naturalizations and the economic and social integration of immigrants. Working Paper 2014:11, Institute for Evaluation of Labour Market and Education Policy (IFAU), 2014.

Christina Gathmann and Julio Garbers. Citizenship and integration. *Labour Economics*, 82:102343, 2023. doi: 10.1016/j.labeco.2023.102343.

Christina Gathmann and Nicolas Keller. Access to citizenship and the economic assimilation of immigrants. *The Economic Journal*, 128(616):3141–3181, 2018. doi: 10.1111/ecoj.12546.

Christina Gathmann and Ole Monscheuer. Naturalization and citizenship: Who benefits? *IZA World of Labor*, 2020(125):125v2, 2020. doi: 10.15185/izawol.125.v2.

Yajna Govind. Is naturalization a passport for better labor market integration? Evidence from a quasi-experimental setting. PSE Working Paper 2021-42, Paris School of Economics, 2021.

Davide Gritti, Filippo Gioachin, Anna Zamberlan, Eleonora Meli, Raffaele Grotti, Paolo Barbieri, and Stefani Scherer. Citizenship acquisition and labour market outcomes among immigrants in Italy: Evidence from linked survey-register data. *International Migration*, 63:e70069, 2025. doi: 10.1111/imig.70069.

Jens Hainmueller, Dominik Hangartner, and Dalston Ward. The effect of citizenship on the long-term earnings of marginalized immigrants: Quasi-experimental evidence from Switzerland. *Science Advances*, 5(12):eaay1610, 2019. doi: 10.1126/sciadv.aay1610.

Jens Hainmueller, Elisa Cascardi, Michael Hotard, Rey Koslowski, Duncan Lawrence, Vasil Yasenov, and David D. Laitin. The socioeconomic returns to citizenship: A randomized controlled trial. *Proceedings of the National Academy of Sciences*, 123(5):e2516178123, 2026. doi: 10.1073/pnas.2516178123.

John E. Hayfron. The economics of Norwegian citizenship. In Pieter Bevelander and Don J. DeVoretz, editors, *The Economics of Citizenship*, pages 91–104. Holger Press, Malmö, 2008.

Joachim Jarreau. Naturalization policy and the economic integration of immigrants: new evidence for France. Working paper, 2020.

Martin Kahanec and Anželika Zaiceva. Labor market outcomes of immigrants and non-citizens in the EU. *International Journal of Manpower*, 30(1-2):97–115, 2009. doi: 10.1108/01437720910948429.

Wonjoong Kim, Yongseung Han, and Myeong Hwan Kim. Determinations of naturalization and its impact on earnings. *European Scientific Journal*, 11(7):209–228, 2015.

Licheng Liu, Ye Wang, and Yiqing Xu. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data, 2024.

Simonetta Longhi, Peter Nijkamp, and Jacques Poot. A meta-analytic assessment of the effect of immigration on wages. *Journal of Economic Surveys*, 19(3):451–477, 2005. doi: 10.1111/j.0950-0804.2005.00255.x.

Francesca Mazzolari. Dual citizenship rights: Do they make more and richer citizens? *Demography*, 46(1):169–191, 2009. doi: 10.1353/dem.0.0038.

National Academies of Sciences, Engineering, and Medicine. *The Integration of Immigrants into American Society*. The National Academies Press, Washington, DC, 2015. doi: 10.17226/21746. Edited by Mary C. Waters and Marisa Gerstein Pineau.

OECD. *Naturalisation: A Passport for the Better Integration of Immigrants?* OECD Publishing, Paris, 2011. doi: 10.1787/9789264099104-en.

Marco Pecoraro and Philippe Wanner. Do dual citizenship rights make more and healthier citizens? Evidence from a quasi-experiment. Annual Conference of the Italian Economic Association, Bologna, 2018.

Ravi Pendakur and Pieter Bevelander. Citizenship, enclaves and earnings: comparing two cool countries. *Citizenship Studies*, 18(3-4):384–407, 2014. doi: 10.1080/13621025.2014.905279.

Floris Peters, Maarten Vink, and Hans Schmeets. Anticipating the citizenship premium: Before and after effects of immigrant naturalisation on employment. *Journal of Ethnic and Migration Studies*, 44(7):1051–1080, 2018. doi: 10.1080/1369183X.2017.1367650.

Floris Peters, Hans Schmeets, and Maarten Vink. Naturalisation and immigrant earnings: Why and to whom citizenship matters. *European Journal of Population*, 36(3):511–545, 2020. doi: 10.1007/s10680-019-09540-1.

Sergiy Pivnenko and Don DeVoretz. The recent economic performance of Ukrainian immigrants in Canada and the US. IZA Discussion Paper 913, IZA, 2003.

Vahan Sargsyan. Treatment-related naturalization premiums in two European countries: Evaluation and comparison. Working Paper 585, CERGE-EI, 2017.

Heidi Shierholz. The effects of citizenship on family income and poverty. Briefing Paper 256, Economic Policy Institute, 2010.

Giacomo Solano and Thomas Huddleston. Migrant integration policy index 2020. Technical report, Barcelona Centre for International Affairs and Migration Policy Group, Barcelona and Brussels, 2020. URL https://www.mipex.eu.

Jan-Jan Soon. Flags and anthems: naturalisation effects on income and employment. *International Journal of Social Economics*, 44(4):491–504, 2017. doi: 10.1108/ijse-06-2015-0179.

T. D. Stanley and Hristos Doucouliagos. Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1):60–78, 2014. doi: 10.1002/jrsm.1095.

Max Friedrich Steinhardt. Does citizenship matter? the economic impact of naturalizations in Germany. *Labour Economics*, 19(6):813–823, 2012. doi: 10.1016/j.labeco.2012.09.001.

Max Friedrich Steinhardt and Jan Wedemeier. The labor market performance of naturalized immigrants in Switzerland—new findings from the Swiss labor force survey. *Journal of International Migration and Integration*, 13(2):223–242, 2012. doi: 10.1007/s12134-011-0213-5.

Eva Vivalt. How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6):3045–3089, 2020. doi: 10.1093/jeea/jvaa019.

Melvin Vooren, Carla Haelermans, Wim Groot, and Henriëtte Maassen van den Brink. The effectiveness of active labor market policies: A meta-analysis. *Journal of Economic Surveys*, 33 (1):125–149, 2019. doi: 10.1111/joes.12269.

Wladimir Zanoni and Ailin He. Citizenship and the economic assimilation of Canadian immigrants. IDB Working Paper Series IDB-WP-01127, Inter-American Development Bank, 2021.

Huiquan Zhou and Sungkyu Lee. Effects of US citizenship on wages of Asian immigrant women. *International Journal of Social Welfare*, 22(4):420–430, 2013. doi: 10.1111/ijsw.12010.