

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Structured interviews of Karelian evacuees were published in a four volume set called 'Siirtokarjalaisten tie' (Anon. 1970-1971). The project was part of an effort to record the lives of the Karelian migrants. Over 300 individuals were trained to conduct these interviews which took place between 1968 and 1970 and an effort was made to locate all people who were evacuated from Karelia during the war. Each entry in the published books lists the name, sex, date of birth, birthplace, occupation, year of marriage, reproductive records (name, sex, and date of birth of all children), membership in various organizations and the years and names of all places where they have lived from birth until the time they were interviewed. If they were married, the name, date of birth, birthplace and occupation of their spouse are also listed. These books were scanned and software was developed (Kaira Core and Natural Language Processing (NLP) software designed for use with the Finnish language) to digitize and extract the records (see Loehr et al (2017) for a detailed description of data extraction methods and the MiKARELIA database). We extracted the name, sex, year of birth, occupation, spouse, spouse's occupation, the names of all places they had lived and the years they moved, their year of marriage and their number of children. Overall there are data on approximately 250,000 individuals, including children and spouses, but here we focus on a subset of 26,757 individuals who were personally interviewed for this project, were born in Karelia, lived in Karelia immediately prior to the Soviet invasion in 1939 and for whom data was complete for all variables of interest.

Data analysis

These data were analyzed using the rethinking package in R version 3.5.1. We used Bayesian inference for all statistical analyses. In a Bayesian framework, each model conditions data on prior probability distributions and uses Monte-Carlo methods to generate posterior distributions for the parameters. The priors are the initial probabilities for each possible value of each parameter. This type of analysis allows us to compare posterior distributions across occupational categories, sexes, age classes, marriage types and migration profiles

without relying on specific post-hoc tests (McElreath 2015) while obviating the need to adjust for multiple comparisons (Gelman, Hill, and Yajima 2012). We are also better able to visualize and interpret differences between parameter estimates relative to a specific value by reporting and displaying the entire posterior distribution for each predictor and showing the highest density intervals (HDI) to reveal the most credible values for each parameter estimate. Here we assume that 95% highest density intervals which do not include zero are evidence that the parameter value is credibly different from the baseline. The R code for all analyses can be found in Supplementary Materials section 'R Code for all Models'.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that supports the findings of this study and code for all the models used to generate these results and produce the figures are available on Github: https://github.com/robertlynch66/NHB_ms_revisions.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is an observational study which provides a quasi natural experiment on some of the factors which influence the reproduction and integration of a population of evacuees during World war II.
Research sample	Structured interviews of Karelian evacuees were published in a four volume set called 'Siirtokarjalaisten tie' (Anon. 1970-1971). The project was part of an effort to record the lives of the Karelian migrants. Over 300 individuals were trained to conduct these interviews which took place between 1968 and 1970 and an effort was made to locate all people who were evacuated from Karelia during the war. Each entry in the published books lists the name, sex, date of birth, birthplace, occupation, year of marriage, reproductive records (name, sex, and date of birth of all children), membership in various organizations and the years and names of all places where they have lived from birth until the time they were interviewed. If they were married, the name, date of birth, birthplace and occupation of their spouse are also listed. These books were scanned and software was developed (Kaira Core and Natural Language Processing (NLP) software designed for use with the Finnish language) to digitize and extract the records (see Loehr et al (2017) for a detailed description of data extraction methods and the MiKARELIA database). We extracted the name, sex, year of birth, occupation, spouse, spouse's occupation, the names of all places they had lived and the years they moved, their year of marriage and their number of children. Overall there are data on approximately 250,000 individuals, including children and spouses, but here we focus on a subset of 26,757 individuals who were personally interviewed for this project, were born in Karelia, lived in Karelia immediately prior to the Soviet invasion in 1939 and for whom data was complete for all variables of interest.
Sampling strategy	This is a population based database and an attempt was made to interview all evacuees who were still alive in 1970. However, because some evacuees were no longer alive in 1970 some of the older evacuees are missing from our sample. Also the sample used was based on individuals for whom complete data was available (e.g sex, age, occupation, spouse, spouses occupation, birth place, spouses birth place, all the places where they lived, moved, and all their children).
Data collection	Each entry in the published books lists the name, sex, date of birth, birthplace, occupation, year of marriage, reproductive records (name, sex, and date of birth of all children), membership in various organizations and the years and names of all places where they have lived from birth until the time they were interviewed. If they were married, the name, date of birth, birthplace and occupation of their spouse are also listed. These books were scanned and software was developed (Kaira Core and Natural Language Processing (NLP) software designed for use with the Finnish language) to digitize and extract the records (see Loehr et al (2017) for a detailed description of data extraction methods and the MiKARELIA database). We extracted the name, sex, year of birth, occupation, spouse, spouse's occupation, the names of all places they had lived and the years they moved, their year of marriage and their number of children.
Timing	Interviews were conducted between 1968 and 1970
Data exclusions	Individuals with missing information on any of the variables of interest -- sex, age, occupation, spouse, spouses occupation, birth place, spouses birth place, all the places where they lived, moved, and all their children-- were excluded from these analyses.

Non-participation	NA
Randomization	Four groups and their combinations were delimited for these analyses in a quasi experimental design. These groups are evacuees who 1) returned to Karelia, 2) those who remained in western Finland, 3) those who were married before the war ended, and 4) those who were married after the war ended.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging