The Uneven Impact of Mobility on the Segregation of Native and Foreign-born Individuals

A Preprint

Yuan Liao*
Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden
yuan.liao@chalmers.se

Jorge Gil
Department of Architecture and Civil Engineering
Chalmers University of Technology
Gothenburg, Sweden
jorge.gil@chalmers.se

Sonia Yeh
Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden
sonia.yeh@chalmers.se

Rafael H. M. Pereira
Institute for Applied Economic Research (Ipea) - Brazil
Data Science Division, Brazil
rafael.pereira@ipea.gov.br

Laura Alessandretti
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Lyngby, Denmark
lauale@dtu.dk

July 25, 2024

Abstract

Segregation is a key challenge in promoting more diverse and inclusive cities. Research based on smartphone data has revealed that segregation can extend beyond residential areas into everyday activities like visiting shops and restaurants. The impact of these activities on segregation, however, is unclear. Some studies suggest that they promote mixing, while others indicate they reinforce segregation. Here, we elucidate how day-to-day mobility shapes overall segregation levels, looking at the distinctive segregation experienced by native and foreign-born individuals. Our study is based on ∼320,000 smartphone trajectories collected in Sweden, where immigration creates profound divides. We find that while mobility levels generally promote mixing for native-born individuals, foreign-born individuals remain segregated in their out-of-home activities. Using counterfactual simulations, we show that this heterogeneous effect of mobility on experienced segregation results mainly from two mechanisms: homophily and limited travel, i.e., foreign-born individuals (i) prefer destinations visited by similar individuals, and (ii) have limited mobility ranges. We show that homophily plays a minor role, while limited mobility, associated with reduced transport access, limits opportunities for foreign-born to diversify their encounters. Our findings reconcile conflicting literature and suggest that enhancing transport accessibility in foreign-born areas could reduce social segregation.

Keywords Mobile phone data · Social segregation · Homophily · Mobility patterns · Transport access

*Corresponding author. Also affiliated with Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark.

1 Introduction

Segregation based on country of origin is a critical problem in many cities worldwide, which has been shown to perpetuate social and economic inequalities [Musterd et al., 2017, Chetty et al., 2022, Sousa and Nicosia, 2022]. The issue is especially acute in countries that have experienced substantial influxes of refugees and migrants in a short period, and has triggered public debates and political attention worldwide [Iceland and Nelson, 2010, Malmberg et al., 2013, Jarvis et al., 2023].

Research in urban segregation has intensified and expanded in the last decade [Liao et al., 2024]. Traditionally, studies have focused on *residential segregation* [Duncan and Duncan, 1955, Feitosa et al., 2007, Barros and Feitosa, 2018], exploring how people from diverse backgrounds distribute in their residence. Recent research enabled by detailed smartphone data has shifted towards examining experienced segregation, which also considers people's exposure to different groups in other spaces, such as workplaces, shops, or leisure areas [Kwan, 2013, Moro et al., 2021, Li et al., 2022]. Understanding *experienced segregation* beyond residential boundaries is critical for developing more effective and nuanced interventions to reduce socio-spatial segregation, compared to altering residential location patterns.

Within this recent body of empirical work, however, findings on how day-to-day activities outside the home influence experienced segregation are mixed. Some research indicates that individuals encounter less segregation outside their area of residence, a phenomenon linked to the fewer constraints faced when choosing daily destinations compared to permanent homes [Park and Kwan, 2018, Garlick et al., 2022, Fuentes et al., 2022, Athey et al., 2021, Moya-Gómez et al., 2021, Shin, 2021, Silm et al., 2021, Xian et al., 2022, Xu et al., 2022, Kronenberger and De Saboya, 2017, Atuesta et al., 2018]. These studies suggest that the flexibility of daily mobility creates a more inclusive and diverse social environment. Conversely, other studies find that mobility outside the home tends to reinforce residential segregation [Hedman et al., 2021, Hilman et al., 2022, Zhang et al., 2022], attributed to two main reasons. The first reason would be homophily, according to which individuals gravitate towards people with similar backgrounds [Xu et al., 2022]. Homophily would manifest as destination preference, driving individuals to visit certain places over others. The second reason would be the participation in daily activities close to home, which could result either from limited transportation access [Vachuska, 2023] or sufficient local services [Abbiasov et al., 2024]. Nonetheless, the effect of these mechanisms, i.e., destination preference and limited travel, on different groups remains unclear, partly due to the difficulties separating these intertwined factors.

Here, we test the hypothesis that day-to-day mobility impacts segregation experiences differently across various groups, i.e., native-born vs. foreign-born, to partly explain the conflicting findings in the literature. Further, we establish the influence of destination preference and mobility range on experienced segregation using statistical techniques in simulated counterfactual scenarios. Our analysis focuses on Sweden, where the issue of segregation by country of birth, i.e., between native-born and foreign-born outside Europe, is particularly critical. Here, the percentage of the foreign-born population has increased from 11% to 21% between 2000 and 2022, making Sweden the country with the second-largest foreign-born population share among all OECD (Organisation for Economic Co-operation and Development) countries [OECD, 2022]. Our study is based on a large-scale smartphone dataset capturing ~ 30 million geolocations for $322,919$ individual devices over seven months in 2019.

2 Results

2.1 Daily mobility reduces segregation overall

Aiming to establish how segregation is affected by daily mobility to participate in out-of-home, we estimate for each individual their (i) *residential segregation* and their (ii) *experienced segregation*. Note that we use post-stratification techniques to mitigate the effect of sampling and population biases in our dataset (see more details in the Methods section, Reducing sampling and population biases).

We define individuals' *residential segregation*, denoted as ICE_r, using the Indicator of Concentration at Extremes (see the Methods section, Segregation indicator) computed within their demographic statistics area of residence (see the Methods section, Stay and home detection, and Socioeconomic attributes assignment). Intuitively, ICE_r reaches its most extreme values for individuals residing solely among foreign-born ($ICE_r = 1$) or native-born ($ICE_r = -1$). A value of $ICE_r = 0$ indicates a residential composition that mirrors the national average in Sweden – comprising 80.4% native-born, 11.1% foreign-born outside Europe, and 8.5% other (foreign-born in Europe). We categorise areas with $|ICE_r| > 0.2$ as segregated, because they deviate significantly from an expected distribution where residency is independent of birth background, at $p < 0.01$ (see the Methods section). Consequently, we classify individuals into three groups: those for which $ICE_r < -0.2$, living in native-born segregated areas (N); those with $ICE_r > 0.2$, living in foreign-born segregated areas (F); and those with $-0.2 \leq ICE_r \leq 0.2$ living in mixed areas (M). We find pronounced residential segregation in Sweden, with 40.8% of individuals living in areas with a higher percentage of native-born than the national average (Group N), 18.4% living in areas with a higher percentage of foreign-born (Group F), and 40.8% in mixed areas (Group M). Segregation is especially marked in Sweden's three largest municipalities – Stockholm, Gothenburg, and Malmö – where over 40%, 49%, and 45% of areas respectively, exhibit a significant average level of residential segregation, $|ICE_r| > 0.2$ (see Fig. 1a, top row).

For each individual, we assess their *experienced segregation*, denoted as ICE_e, by measuring the composition of the people they encounter during activities outside the home on non-holiday weekdays (for further details, see the Methods section, Residential and experienced segregation). The focus on non-holiday weekdays narrows the scope of our analysis to understand the role of routine day-to-day activities. Encounters are defined as co-locations within the same 30-minute interval at the same area with an average size of $0.087km^2$. Also, in this case, ICE_e is measured as the Indicator of Concentration at Extremes: $ICE_e = 0$ indicates that the demographic composition of the encounters aligns with the overall national composition, while extreme values $ICE_e = 1$ and $ICE_e = -1$ occur if encounters are exclusively with either foreign-born or native-born individuals, respectively. Experienced segregation, although correlated with residential segregation (corr. coefficient: 0.48), is generally lower (see Fig. 1b). Specifically, ICE_e is lower than ICE_r for 62% of individuals in our dataset. We find that 28%, 40%, and 73% of areas in Stockholm, Gothenburg, and Malmö exhibit high experienced segregation levels, $|ICE_e| > 0.2$. Interestingly, in contrast to the patterns observed for residential segregation (see Fig. 1a, top row), ICE_e is always negative in these cities (see Fig. 1a, bottom row). This suggests that mobility may affect the native and foreign-born groups differently.

2.2 The uneven impact of daily mobility on segregation

Figure 1c illustrates the shift in segregation distribution – from residential to outside home experiences – for each group. We find that for the native-born individuals (Group N), the average segregation becomes significantly lower from 0.385 ± 0.001 to -0.053 ± 0.001 (standard errors of the median, applied similarly in the following sections) once we account for daily mobility in the analysis. The reduced segregation reveals that, on average, native-born individuals are exposed to more social diversity in their daily activities, where they experience less segregation than in their residential areas. Conversely, for the foreign-born individuals (Group F), the impact of mobility on segregation is limited, resulting in a modest reduction with average values shifting from -0.380 ± 0.002 to -0.295 ± 0.001.

To assess the impact of mobility on individual segregation (ICE_e vs. ICE_r), we categorize individuals into three groups. This classification mirrors the previously employed residential segregation methodology (see the Methods section). Specifically, individuals are grouped based on their ICE_e as follows: those with $ICE_e > 0.2$ are categorised as native-born segregated, those with $ICE_e < -0.2$ as foreign-born segregated, and those within the range $-0.2 \leq ICE_e \leq 0.2$ as exposed to a mixed group of individuals (see Fig. 1d). Our analysis reveals notable shifts between residential and experienced segregation (see Fig. 1d). Among individuals initially identified as native-born (Group N) based on residential data, 76% transition into the mixed category (Group M) when mobility is considered. Conversely, only 38% of individuals initially identified as foreign-born (Group F) transition to the mixed category. The remaining 62% of individuals in this group continue to experience a strong and significant level of foreign-born segregation during their daily activities (see Table B.3-Scenario Empirical).

2.3 The marginal role of destination preference

We investigate why foreign-born individuals experience high segregation during their daily activities. We consider four key factors that influence segregation [Liao et al., 2024] (see Fig. 2a): (i) *mobility range*: the extent to which one travels beyond his/her local area; (ii) *destination preference*: individuals' inclination to visit certain places due to homophily; (iii) *urban structure*: the geographical distribution of residences, amenities and transportation networks; and (iv) *lifestyle*: the type and timing of individuals' daily activities. In this study, we specifically focus on estimating the effects of destination preference and mobility range based on a counterfactual analysis through simulations. We do not concentrate on lifestyle choices because we find they have a limited impact on segregation in our dataset. In particular, we observe consistent activity levels and similar distribution patterns across types of locations among native-born and foreign-born individuals (see Section B.2). On the other hand, we do not extensively examine the role of urban structure, which sets a pre-condition for daily activities but is hardest to alter by individual decisions.

Initially, we test the hypothesis H_h that experienced segregation in the foreign-born group is fully explained by destination preference (homophily). To this end, we perform simulations that maintain the real distribution of amenities and residences, lifestyle choices, and mobility ranges but where visit patterns are altered. In this *no-dest. preference* simulation, for each recorded trip in our dataset, we substitute the original destination with a random nearby location (within 1 km) of the same type (e.g., bar, restaurant, café, see more details in the Methods section). To confirm or reject H_h, we measure ICE_e in the counterfactual scenario, *no-dest. preference*. Significant values of ICE_e would lead to

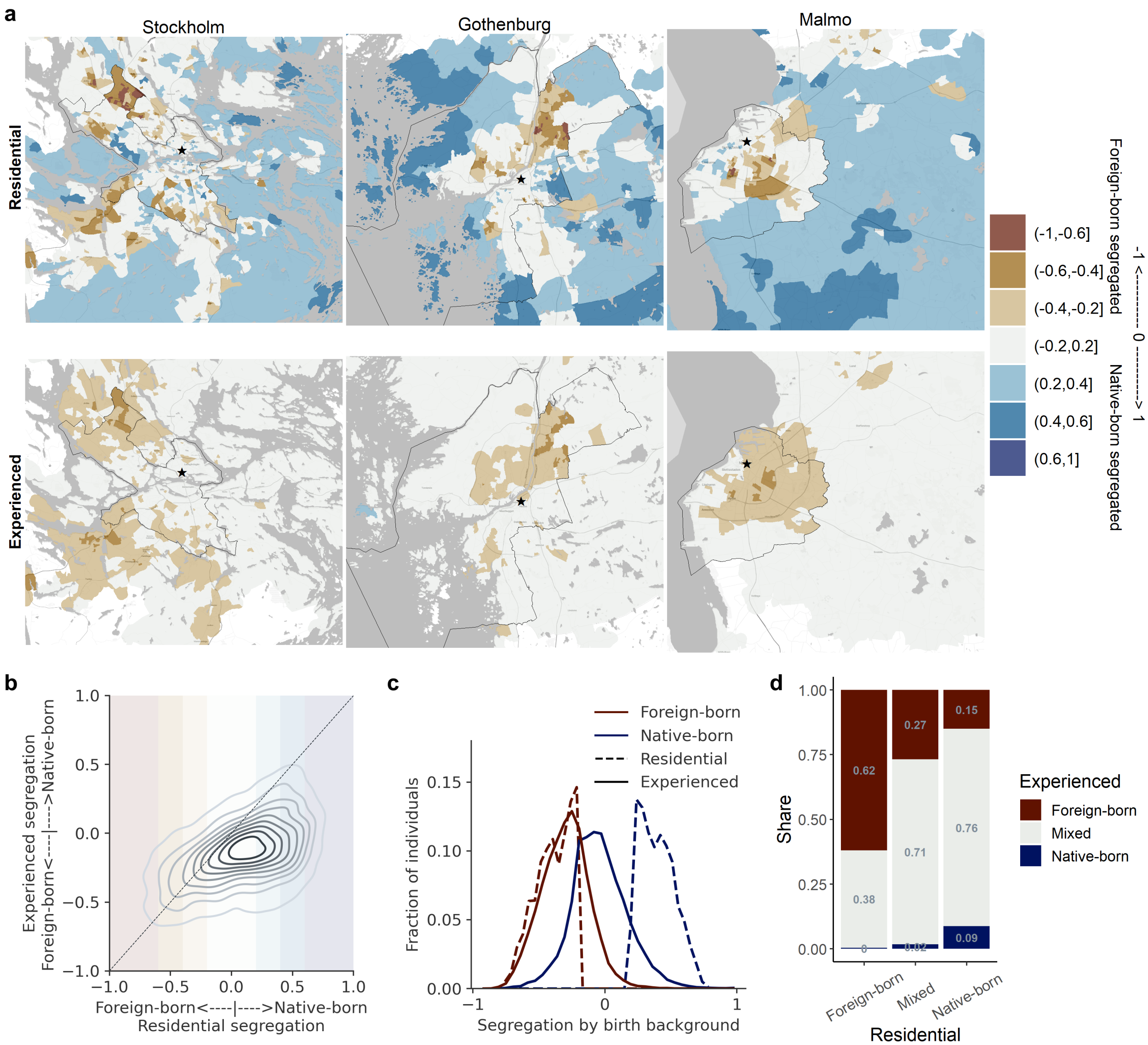

Figure 1: **Residential segregation and experienced segregation outside residential area**. **a**, Choropleth maps showing residential (top row, ICE_r) and experienced (bottom row, ICE_e) segregation in Sweden's three major cities: Stockholm, Gothenburg, and Malmö (left to right). Colors capture the average segregation within each statistical area. Black lines show municipality boundaries. **b**, Scatter density plot of Experienced (ICE_e) vs Residential (ICE_r) segregation for all individuals in the dataset. Background colors indicate levels of residential segregation (see color bar in subplot a). The gray contour lines represent areas with similar data point densities, as determined by a kernel density estimate (KDE) plot. The diagonal line indicates where $ICE_e = ICE_r$. **c**, The distribution of residential (dashed lines) and experienced (full lines) segregation for individuals in the foreign-born (red) and native-born (blue) segregated groups. **d**, Share of individuals by experienced segregation (ICE_e, y-axis) across different residential segregation groups (ICE_r, x-axis). Individuals are categorized by residential segregation into three groups: $ICE_r < -0.2$ (foreign-born), $ICE_r > 0.2$ (native-born), and $-0.2 \leq ICE_r \leq 0.2$ (mixed). The same categorization criteria are applied to their experienced segregation along the y-axis.

the rejection of our hypothesis. To assess the significance of ICE_e we perform significance tests against a Random mixing scenario (see Methods, Identifying segregated individuals).

Indeed, our analysis leads us to reject the hypothesis H_h. Under the *no-dest. preference* scenario, we observe an average reduction in absolute experienced segregation $|ICE_e|$ for 71% of the individuals in the foreign-born group (F). Interestingly, while this reduction is statistically significant, its magnitude suggests having a small effect size according to Cohen's test (see Table 1). Under the *no-dest. preference* scenario, in fact, we still observe significant segregation for the foreign-born group, characterized by $ICE_e = -0.234 \pm 0.001$ (Fig. 2b and Table B.3). For comparison, under the *no-dest. preference* scenario, the native-born group (N) also experiences a small reduction in absolute experienced segregation (58% of individuals) and continues to present no significant segregation with $ICE_e = -0.061 \pm 0.001$ (Table B.3). Overall, the analysis reveals that, when presented with two similar venues in close proximity, individuals from both Groups N and F exhibit a slight preference for visiting venues frequented by people similar to themselves. This tendency is marginally more pronounced among Group F. However, this effect alone does not fully explain the high degree of experienced segregation observed in empirical data, with statistical test results detailed in Section B.5 (Table B.3 and Fig. B.4).

2.4 The key role of mobility range

We proceed to test the hypothesis H_{ht} that *destination preference* and *mobility range*, when combined, can explain the observed levels of segregation in the foreign-born group.

To evaluate this hypothesis, we conduct an *equalized mobility & no-dest. preference* randomization simulation aimed at standardizing individuals' mobility range and minimizing their destination preferences (detailed in the Methods section). For each trip made by an individual in the dataset, we replace the destination with a random venue of the same type located at a distance d (with a buffer of ± 30 m) from the individual's home. To ensure that all individuals have the same mobility range, we extract the distance d as a probability distribution from the average empirical distribution of distances between individuals' homes and visited locations (see the Methods section). As in the previous section, to confirm or reject H_{ht}, we measure ICE_e in the counterfactual scenario, where significant values of ICE_e change would lead to rejecting H_{ht}.

We find that, under the counterfactual scenario, the average segregation for Group F is significantly reduced (see Table 1) to a non-significant level of $ICE_e = -0.066 \pm 0.0003$ (Fig. 2b and Table B.3). Hence, we cannot reject the hypothesis H_{ht}. Our result suggests that when combined, mobility range and destination preference can explain most of the level of experienced segregation observed in the foreign-born group. For comparison, under this counterfactual scenario, Group N also experiences a small reduction in absolute segregation, and continues to show no significant levels of segregation with $ICE_e = -0.049 \pm 0.0003$ (Table B.3). Overall, the analysis reveals that limited mobility range plays a key role in the experienced segregation of foreign-born individuals. The corresponding statistical test results are detailed in Supplementary Section B.5 (Table B.3 and Fig. B.4).

Table 1: **The effects of destination preference and mobility range on segregation**. The table shows the results of the statistical tests measuring the difference between empirical data and the two counterfactual scenarios for the native-born (N) and foreign-born (F) segregated groups. Diff. is the statistic of the Weighted Mann-Whitney U test between simulated and empirical distributions. All values are statistically significant, with $p < 0.001$. The effect size is estimated based on Cohen's d, the magnitude of the difference between the two distributions. A Cohen's d value around 0.2 is typically considered a small effect. A value around 0.5 is considered a medium effect. A value around 0.8 or larger is considered a large effect.

Effect	Group	Diff.	Cohen's d	Effect size
Destination preference	N	0.012	-0.06	Very small
	F	-0.10	0.34	Small
Destination preference & mobility range	N	-0.02	-0.01	Very small
	F	-0.41	1.81	Large

2.5 Mobility range and destination preference shape exposure between groups

To better understand the factors driving segregation, we study how individuals from different groups are exposed to each other during their daily activities. Specifically, we calculate the proportion of potential encounters each individual has with members of Groups N, F, and M (Fig. 2c). In a scenario of random mixing, these proportions would reflect the distribution of individuals in each group: 42.3% in Group N, 17.6% in Group F, and 40.1% in Group M. Therefore, for each individual, we compute the deviation from these baseline values (Fig. 2d). The significance of such deviations

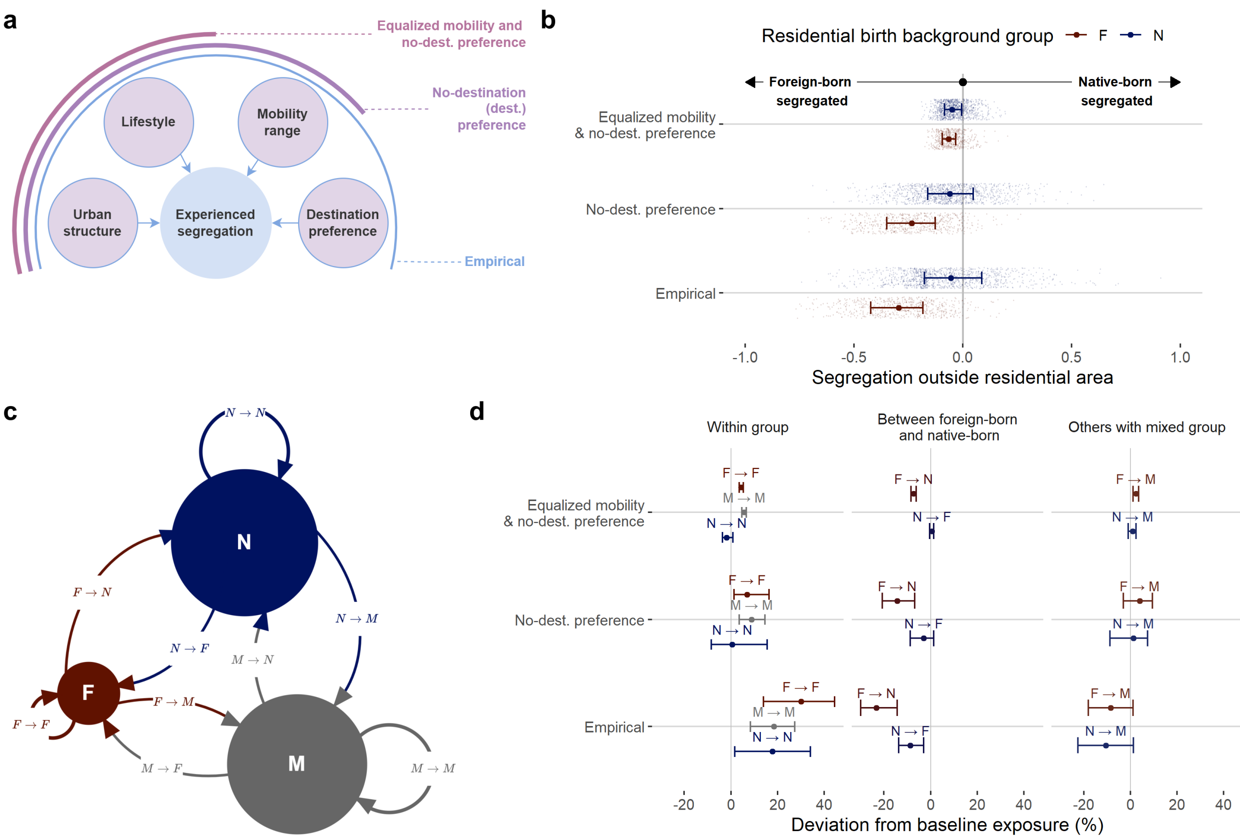

Figure 2: **The effects of destination preference and mobility range on experienced segregation**. **a**, Empirical experienced segregation is impacted by urban structure, individuals' lifestyle, mobility range, and destination preference. The first counterfactual scenario minimizes the effect of destination preference, and the second equalizes mobility range across users. **b**, segregation (ICE_e) experienced by foreign-born (red dots) and native-born (blue dots) individuals in the empirical (bottom), first scenario (mid) and second scenario (top) conditions. Lines show the ranges between the 25th and 75th percentiles. **c**, Schematic network displaying exposure among individuals in foreign-born concentrated areas (Group F, red), native-born concentrated areas (Group N, blue), and mixed group (Group M, gray). Nodes correspond to groups, and their radii are proportional to the share of individuals in the data. The notation $X \to Y$ indicates the typical fraction of an individual's encounters with individuals from Group Y, given that the individual is from Group X. **d**, Group exposure (see subplot c for the notation) in the empirical (bottom), first scenario (mid), and second scenario (top) conditions. Values are shown as the deviation from a baseline capturing expected exposure under random mixing conditions: 42.3%, 17.6%, and 40.1% for exposure with Groups N, F, and M, respectively. Lines represent the ranges between the 25th and 75th percentiles.

is assessed by performing significance tests against a Random mixing scenario (see Methods, Identifying segregated individuals). In our notation, $X \rightarrow Y$ indicates the typical fraction of encounters an individual from Group X has with individuals from Group Y.

Our findings reveal that individuals tend to be exposed to others from their own group than expected. This effect is significant for individuals in Group F (detailed in Supplementary Section B.5, Table B.4 and Fig. B.5). Meanwhile, the other groups only present a slight tendency to have inner-group exposure that is not significant. The proportions of $F \rightarrow F$, $N \rightarrow N$, and $M \rightarrow M$ encounters are $30.0 \pm 0.17\%$, $17.7 \pm 0.14\%$, and $18.4 \pm 0.06\%$ higher, respectively, than what would be expected under random mixing conditions. Conversely, exposure across groups, $F \rightarrow N$ and $N \rightarrow F$ are $-23.3 \pm 0.08\%$ and $-8.7 \pm 0.05\%$ lower than expected.

In the *no-dest. preference* scenario, Group N experiences encounters consistent with baseline (Table B.4), with $N \rightarrow N$ being only $0.6 \pm 0.1\%$ higher and $N \rightarrow F$ only $-3.1 \pm 0.02\%$ lower than expected. Group F continues to experience a slightly unbalanced set of encounters, with $F \rightarrow F$ being $6.9 \pm 0.07\%$ higher and $F \rightarrow N$ being $-14.3 \pm 0.08\%$ lower than expected. We note, however, that this result is not statistically significant (Table B.4), in apparent discrepancy with the results obtained in the previous section for the ICE_e. This difference can be attributed to two factors. First, the categorical group exposure considers only three groups of individuals, providing less detail than the continuous quantification of the experienced segregation level. Second, although Group F's exposures to Group F and Group N are insignificantly different from random mixing in the *no-destination preference* scenario, the combined deviations result in Group F remaining largely segregated overall.

Finally, in the *equalized mobility & no-dest. preference* scenario, Group N remains close to the baseline exposure to its own group ($N \rightarrow N$, $-1.8 \pm 0.02\%$) and Group F ($N \rightarrow F$, $0.6 \pm 0.01\%$). Similarly, in this scenario, the level of interaction of individuals in Group F with members of their own group ($F \rightarrow F$, $4.3 \pm 0.01\%$), and with Group N ($F \rightarrow N$, $-7.3 \pm 0.01\%$) do not deviate significantly from the baseline (see Table B.4).

In summary, these results show that destination preference has a minor but significant influence on the group exposure of the foreign-born segregated population. After minimizing the destination preference, there is no significant deviation from the baseline for Group F. Further, equalizing the mobility range slightly narrows the gaps between Group F and the others. For their native-born counterparts, these two factors have a minimal effect due to the already low level of experienced segregation in their daily activities outside the home.

2.6 Reducing segregation by facilitating mobility

The results in the previous sections suggest that mobility range is a key factor driving differences in the experienced segregation of foreign-born and native-born individuals. To better understand this effect, we dig into differences in mobility behavior between the two groups.

Compared to Group F, individuals in Group N tend to travel longer distances to reach their day-to-day destinations (see Fig. 3a and b). Interestingly, for both groups, the experienced segregation is higher in venues closer to home, such as education or religious destinations (see Fig. 3a). Instead, experienced segregation is lower in venues located further from home, such as retail and financial venues (see Fig. 3a).

Within each group, the level of experienced segregation is associated with the radius of gyration [Xu et al., 2018], capturing the extent of individuals' whereabouts (see Fig. 3b). Individuals with larger mobility ranges tend to experience lower levels of segregation, especially for Group F (corr. coefficient = -0.21, $p < 0.001$).

To understand the role of transport accessibility in segregation, we investigate their relationship by focusing on individuals with low car ownership. This specific group is assumed to rely primarily on public transportation and active modes of travel. We observe that, within this group, individuals in Group F have much lower job accessibility by transit than those in Group N (see Fig. 3c). In Group F, higher job accessibility by public transit is associated with less experienced segregation outside residential areas (corr. coefficient = -0.16, $p < 0.001$).

Our results reveal that longer mobility ranges and better transport accessibility are associated with reduced segregation.

3 Discussion

In this study, we used mobile phone GPS records to examine the division between native-born and foreign-born (outside Europe) populations in Sweden. We found that the segregation experiences of foreign-born and native-born individuals are affected differently by their mobility patterns when engaging in out-of-home activities. Native-born individuals experience non-significant levels of segregation while performing routine activities outside the home. In contrast, residents of foreign-born segregated areas experience less, but still significant, levels of segregation during

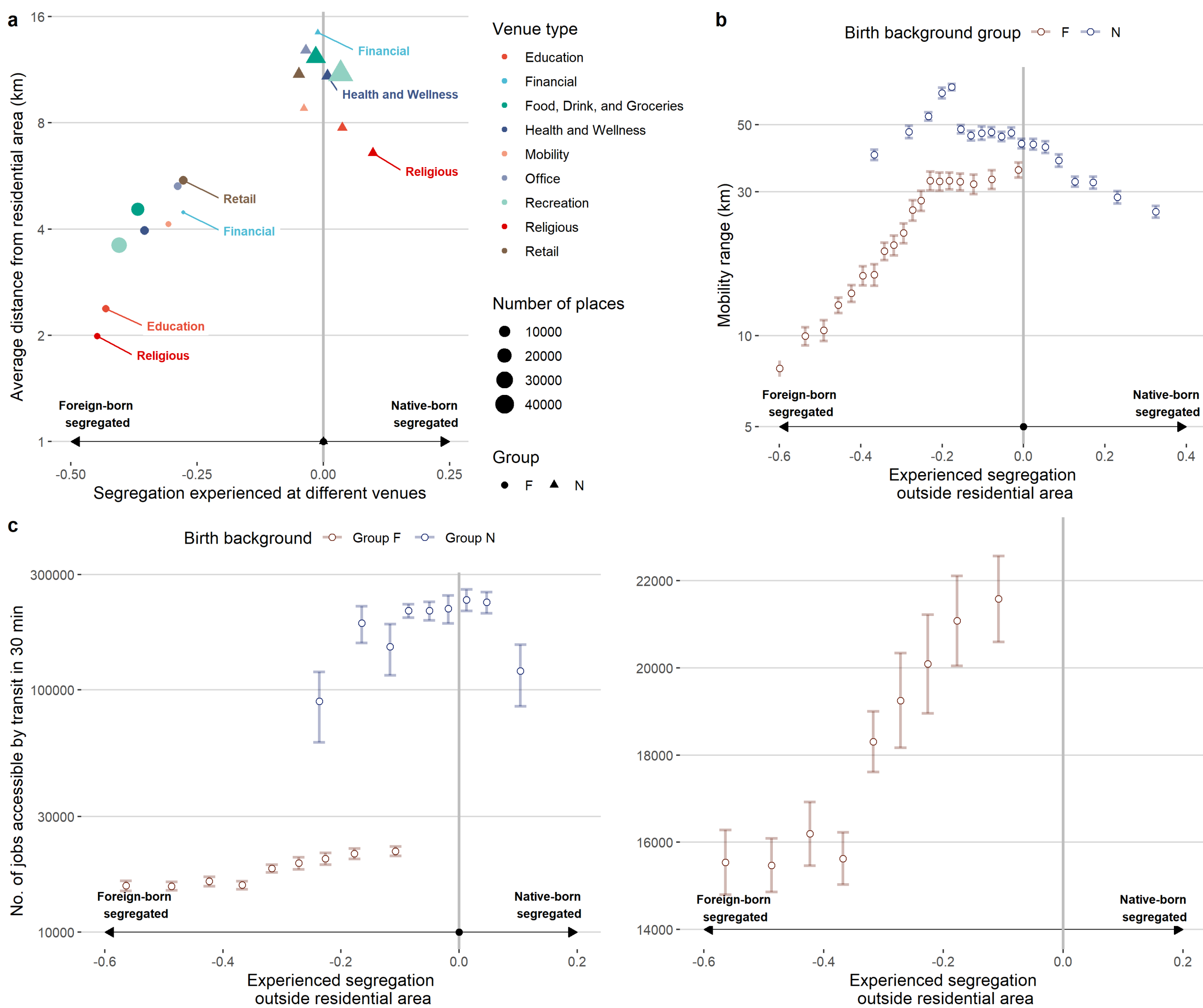

Figure 3: **Segregation is associated with low mobility range and transport accessibility**. **a**, The average segregation experienced in visited destinations of different types against their average distance from visitors' homes. Results are shown for foreign-born (circles) and native-born (triangles) segregated individuals. **b**, Mobility range of individuals by their experienced segregation. The mobility range is represented by the radius of gyration (km). Lines show the median errors. Red lines/dots are the results for those living in foreign-born segregated areas (Group F), and blue lines/dots are for native-born segregated areas (Group N), divided into 19 groups based on the quantile breaks of their respective segregation level. **c**, Transit accessibility of individuals with low car ownership by their experienced segregation, divided into 9 groups based on the quantile breaks of their respective segregation level. The right part is a zoom-in of Group F. Transit accessibility is measured by the count of jobs reachable within 30 minutes by public transport from one's home. Low car ownership is identified as ownership rates under the 25th percentile, i.e., below 0.28 per capita.

their outside-home routine activities, because they tend to visit locations popular to other foreign-born individuals. Through counterfactual simulations, we tested two hypotheses to understand the factors driving this heterogeneous effect. First, we explored whether foreign-born individuals prefer destinations frequented by those with similar backgrounds (homophily hypothesis). We observed that destination preference is only marginally responsible for the segregation levels experienced by foreign-born individuals. Secondly, we tested whether the observed heterogeneous effect of mobility patterns on segregation results due to differences in mobility ranges between foreign-born and native-born individuals. Our findings indicate that mobility ranges play a crucial role, with foreign-born individuals travelling shorter distances around their residential areas compared to native-born. Through counterfactual simulations, we demonstrated that if everyone had the same mobility range and no destination preference, the segregation experienced by foreign-born individuals would be negligible. Importantly, we found that limited mobility ranges among foreign-born individuals are associated with lower car ownership and reduced transport accessibility.

Our analysis contribute to the recent stream of work on mobility and segregation through the analysis of GPS trajectories collected from smartphones [Liao et al., 2024]. Our findings suggest that, in Sweden, day-to-day activities overall reduce experienced segregation by country of origin, similarly to several previous studies [Park and Kwan, 2018, Garlick et al., 2022, Fuentes et al., 2022, Athey et al., 2021, Moya-Gómez et al., 2021, Shin, 2021, Silm et al., 2021, Xian et al., 2022, Xu et al., 2022, Kronenberger and De Saboya, 2017, Atuesta et al., 2018]. They align with previous studies, showing that mobility range relates to individuals' segregation experiences and that activities/locations close to individual residences are more segregated [Moro et al., 2021]. Additionally, visits to venues in the city centre, such as retail locations [Nilforoshan et al., 2023], are associated with lower levels of experienced segregation. Previous studies showed that travelling outside one's residence generally reduces the level of experienced segregation [Östh et al., 2018, Abbiasov et al., 2024].

It is important to acknowledge some limitations in our analysis. First, similar to previous research [Moro et al., 2021, Nilforoshan et al., 2023], we estimate individuals' socioeconomic attributes based on census data from their areas of residence. Utilizing data that integrates both socioeconomic attributes and mobility information would improve the accuracy of our analysis. Secondly, our counterfactual scenarios treat broad categories of locations (e.g., restaurants, cafés) as homogenous, overlooking differences in cost, attractiveness, and other characteristics. Third, the observed relationship between transport access and segregation is correlational. Future studies using longitudinal data could explore whether improved transport access facilitates greater interaction between foreign-born and native-born individuals. Finally, as noted in previous studies [Schnell and Haj-Yahya, 2014, Zhou and Cheng, 2019, Dorman et al., 2020], mere co-presence or exposure in a given area does not always equate to meaningful social interactions. Further research is needed to understand how co-presence translates into meaningful social integration, employing alternative methods.

Overall, our work suggests that improving transport accessibility could help reduce social segregation [Huang et al., 2022]. Enhanced accessibility broadens individuals' freedom to reach opportunities across employment, healthcare, and education services [Pereira et al., 2017]. Prioritizing accessible and affordable transport solutions, particularly in areas with high foreign-born populations, could promote social mixing and create more inclusive communities.

4 Methods

4.1 Data preparation and segregation computation

4.1.1 Mobility data

This study uses anonymized mobile application data from GPS records collected through location-enabled smartphone applications. The dataset includes individuals who consented to anonymously share their data through a General Data Protection Regulation-compliant framework. In 2020, data was shared under a contract with Pickwell, who provided access to de-identified and privacy-enhanced mobility data. All researchers were required not to share the data further or attempt to re-identify it.

The dataset covers seven months in 2019 (June-Dec) with about 25 million daily GPS time records from 1 million devices residing in Sweden, giving a total of approximately 5.2 billion records, i.e., 5,250 records per device on average. Assuming that a device is a person, the population covered by this dataset is equivalent to c.a. 10% of the Swedish residents. After further processing, we focus on individuals who reside in Sweden, have sufficient data records, and have reliably identified home zones (see Supplementary Section A.2.1). More details of the dataset can be found in Supplementary Section A.1 Mobility data.

4.1.2 Stay and home detection

To measure experienced social segregation, we need to measure when and where people are engaged in activities. To do so, the first step is to detect *stays* from individual geolocation records (*id, lat, lon, time*). A stay is defined as an instance in which an individual spends a significant amount of time within the same area. These *stay* observations are used for further analysis to characterize individuals' travel and mobility segregation patterns. There are several algorithms for detecting stays [Aslak and Alessandretti, 2020, Hariharan and Toyama, 2004]. In this study, we use the Infostop algorithm [Aslak and Alessandretti, 2020] due to its robustness against measurement noise, scalability to large datasets, and capability of doing multi-user analysis simultaneously. The details, such as the selection of the algorithm parameters, are specified in Section A.2 Stay detection.

Once stays are detected, we identify individuals' residential areas in order to match them with socioeconomic attributes from census data. Residential areas are population grids or census zones that correspond to the most probable location of residence for an individual device. Following previous studies [Moro et al., 2021], we identify residences by first identifying the top 3 most visited places during non-holiday time[2], and then select the most visited place between 10:00 p.m. and 6:00 a.m. In this step, we filter out those individuals without reliable home areas or sufficient data. More details can be found in Section A.2.1 Data filtering.

4.1.3 Socioeconomic attributes assignment

We collect census data from the statistics agency in Sweden [Statistikmyndigheten SCB, 2023]. Data come with two systems of spatial units: Demographic statistics areas (DeSO zones) and population grids with sizes of 250 m or 1000 m (Section A.3). Census data includes information on the following characteristics of each spatial unit (grid and DeSO zone) that are used in this study: total population, birth background (Sweden, outside Sweden in Europe, outside Europe, and others), and job count (daytime job register). These attributes are assigned to each individual device according to the census data and where a given device resides. In this study, we use DeSo zones to calculate residential segregation, and population grids to summarize individuals' experienced segregation and for transport accessibility computation. Section A.3 Census data contains more details of the data used.

4.1.4 Reducing sampling and population biases

Mobile phone application data records the population's geolocations when they use certain applications. Therefore, we have more recorded locations when people are more likely to use their phones, i.e., in the afternoon, evening, and night, compared with other times. We design stay weight (W_{pr} for record r of individual p) to reduce the sampling bias introduced by imbalanced knowledge of individuals' geolocations due to this passive data collection. Specifically, we assign greater significance to time intervals during which fewer locations are registered and lesser significance to those with more recorded locations. This approach helps reduce the sampling bias toward activities at certain times of the day when we have more records. For more details on the stay weight design, please see A.5 Sampling debiasing.

Another bias of mobility data stems from who uses these mobile devices, i.e., how representative the phone users are compared with the actual population. This study assigns weights to individual mobile devices based on their home zones. It uses Inverse Probability Weighting (IPW) to give more weight to less densely populated areas, reducing bias from high device concentration. Extreme weights in sparsely populated areas are controlled using weight trimming. This ensures balanced weights, enhancing the accuracy and reliability of forthcoming statistical analyses. The created individual weight has a notation W_p for a given individual p. The weight reduces the population bias, where we have proportionally more devices in big cities than in the other parts of Sweden, providing a better representation of the population. A.6 Population debiasing contains more details of the population weight design.

At the end of this process, we have a data set of stay records representative of individuals residing in Sweden, which can be used to calculate their social segregation levels.

4.1.5 Segregation indicator

We measure the social segregation between foreign-born (outside of Europe) and native-born populations using a modified version of the Indicator of Concentration at Extremes (ICE) [Massey, 2001], which can be calculated using Equation 1 for a given zone j. This indicator measures how the areas deviate from their national average composition of birth background, which offers a comprehensive and actionable understanding of segregation between native-born and foreign-born. It places local segregation patterns in a national context, aiding in informed decision-making and policy development.

[2]Holiday time: 2019-06-23–2019-08-11 and 2019-12-22–2020-01-01. These times are when most people take summer and winter vacations in Sweden.

$$ICE_j = \frac{\frac{N_j}{w_N} - \frac{F_j}{w_F}}{\frac{N_j}{w_N} + \frac{F_j}{w_F} + \frac{O_j}{w_O}} \tag{1}$$

Here, F_j is the number of foreign-born (outside of Europe) individuals, N_j the number of native-born individuals, and O_j the remaining individuals. The overall fraction of native-born, foreign-born outside of Europe and other populations at the national level are w_N, w_F, and w_O, respectively. We use the data from 2019 when $w_N = 0.804$, $w_F = 0.111$, and $w_O = 0.085$. It is worth noting that the original formula of ICE does not include population weights. We adopt the adjusted ICE using weights because we focus on the relative value of birth background segregation. The comparison between our indicator and the original one is detailed in Supplementary Section B.1 Adjusted ICE vs. its original form.

ICE_j ranges between -1 and 1. In a location with $ICE_j = 0$, the composition of subpopulations equals the national average, $ICE_j = 1$ means that there are only native-born living (or being active) in zone j and $ICE_j = -1$ indicates only foreign-born (outside of Europe).

4.1.6 Residential and experienced segregation

Residential segregation ICE_r is calculated using Census data following Equation 1, with F_j and N_j the foreign-born (outside of Europe) and native-born residents in a census zone j.

Individuals' mobility time history tells us the places they visit, when, and for how long, and the segregation levels at those places describe the individually experienced segregation. To calculate **experienced segregation**, we need to know the visiting segregation of all the analysis zones (detailed in Supplementary Section A.4) an individual visited and aggregate them. Therefore, we first quantify the visiting segregation for various zones across different times. We use the mobility data to find all individuals co-present in each spatiotemporal unit, i.e., which analysis zone (see Supplementary Section A.4) and time interval, where a day is divided into 48 half-hour intervals starting from 00:00 (i=1,2,...,48). We also define whether a day is a weekday or a weekend/holiday. As a result, we have $2 \times 2 \times 48 = 192$ temporal units, considering 48 time intervals of a day, weekday/weekend, and holiday/non-holiday.

Visiting segregation of a given analysis zone j at a temporal unit $(i, weekday, holiday)$ is calculated using the co-present visitors following Equation 1. Unlike the residential segregation measuring residents, we need to calculate F and N for visitors. Each visitor has these attributes from their residential population grid's statistics (Census data). We calculate the weighted sum of these attributes using their population weight to get a more representative picture of the composition of the co-present visitors and, therefore, the level of visiting segregation $ICE_v\left(i, j, weekday, holiday\right)$.

This study continues with the aggregate picture of individually experienced segregation ICE_e in daily mobility to simplify the analysis ($weekday = 1$, $holiday = 0$). For each time interval i, an individual may have stay records in a few different zones j. Therefore, we use the median value of $ICE_v\left(j\right)$ to represent the average experienced segregation at a given temporal unit $ICE_e\left(i\right)$. We further simplify the analysis by calculating ICE_e for an average non-holiday weekday as the mean value of $ICE_e\left(i\right), i = 1, 2, ..., 48$.

At the end of this calculation, each individual has one residential segregation level ICE_r depending on where they reside and one experienced segregation level ICE_e derived from their mobility outside residential areas.

4.2 Explaining birth background segregation

4.2.1 Identifying segregated individuals

Based on the residential segregation level of individuals, we create three groups: those segregated towards foreign-born (Group F), those segregated towards native-born (Group N), and the rest with insignificant segregation (Group M). We conduct a residential randomization simulation to establish thresholds for identifying segregated individuals. This simulation emulates a scenario where individuals in Sweden make random housing/visiting choices, leading to a diverse mix of foreign-born, native-born, and others. We call this scenario Random mixing, which serves as the comparison baseline. In this process, we randomly reassign residences among individuals and then assess the level of residential segregation for each person. After completing this randomization process 100 times, we compile and analyze the distribution of residential segregation levels for these individuals (see Fig. 4).

Residential segregation values falling outside the 99% confidence interval are deemed statistically segregated. Consequently, thresholds of -0.2 and 0.2 are established for classifying individuals. Specifically, those with a residential segregation level below -0.2 are categorized into Group F (foreign-born concentrated), those above 0.2 into Group N (native-born concentrated), and those between -0.2 and 0.2 are placed in Group M (insignificantly segregated).

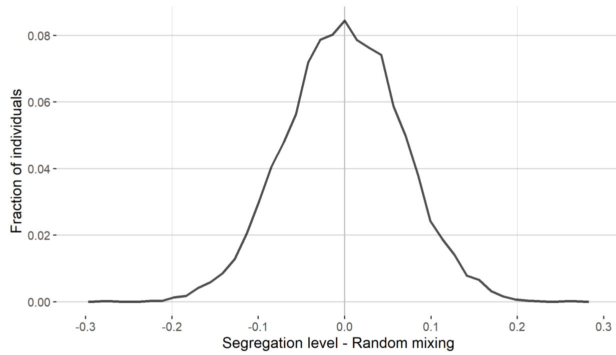

Figure 4: Distribution of residential segregation levels of individuals in a simulated scenario where individuals randomly choose where to live in Sweden. The distribution contains 100 repetitions of the simulation. The data between the two vertical lines correspond to the 99% confidence interval around 0.

Each individual is also categorized into one of the same three groups based on their experienced segregation level sequence at each half-hour interval during an average weekday, $ICE_e(i), i = 1, 2, ..., 48$. If the sequence is significantly greater than 0.2, the individual is classified into Group N. Conversely, if the sequence is significantly less than -0.2, they are placed in Group F. In all other scenarios, the individual falls into Group M. For sequences forming a normal distribution, we apply a one-sample t-test; for non-normal distributions, we utilize the Wilcoxon test.

In the Random mixing scenario, each individual is exposed to others outside their home with the following composition: 42.3% from Group N, 17.6% from Group F, and 40.1% from Group M, reflecting the distribution of these groups in our dataset. To determine significant deviations from this random mixing, we design a procedure to obtain confidence intervals for the share of encounters outside the home. Following the randomization process, we identify the 95% confidence intervals for the presence of the three groups in each individual's encounters. Specifically, any share outside the ranges (-19%, 19%) for Groups N and M, or (-13%, 13%) for Group F, is considered a significant deviation from random mixing. For the group exposure distribution, we utilize the Wilcoxon test to see if it falls outside its 95% confidence interval. If the group exposure distribution is significantly greater than the upper bound or less than the lower bound, it significantly deviates from the Random mixing exposure. In all other scenarios, there is no significant deviation from the Random mixing. The details of the above statistical results can be found in Section B.5.

4.2.2 No-destination preference simulation

The destination preference hypothesis refers to individuals who tend to visit places where there are people similar to them. The No-destination preference simulation aims to reduce the impact of individual preference on the visited places by randomly shifting them to nearby places with similar functions.

We first assign each non-home stay to a point of interest, i.e., the nearest POI within a search radius of 300 m. Then, we randomly shift the non-home stays to a nearby POI with the same category or similar kind. This shifted stay is randomly selected from the POIs between the search radii of 30 m and 1000 m around the original stay record. The details of POI data and how the random shifting is implemented are specified in Section A.7.1 Point of interest data.

For each individual, we ran this simulation 50 times. Finally, we repeat the same calculation described in Section 4.1.6 to get the simulated values of experienced segregation.

4.2.3 Equalized mobility & no-destination preference simulation

We hypothesize that individuals who are reluctant or face challenges in traveling contribute to restricted social mixing with other population groups due to their limited mobility range. To quantify the contribution of mobility range in segregation experiences, we design the Equalized mobility & no-destination preference simulation that equalizes travel distance disparities among individuals in our dataset by assuming a uniform mobility range across them.

We begin by documenting all the locations individuals visited and calculating the distances from their home locations. Focusing on distances less than 1,500 km, we establish bins with 1 km intervals to compute the frequency of visits within each distance range. The resulting distributions, illustrated in Figure 5, encompass all individuals, those residing in areas with foreign-born populations and those with native-born populations. This simulation assumes one travels as an average person, following the distance-decay trend (the gray distribution in Fig. 5). In this case, the individuals in

Group F will increase their mobility range, while Group N will slightly decrease their mobility range compared to their respective empirical mobility patterns.

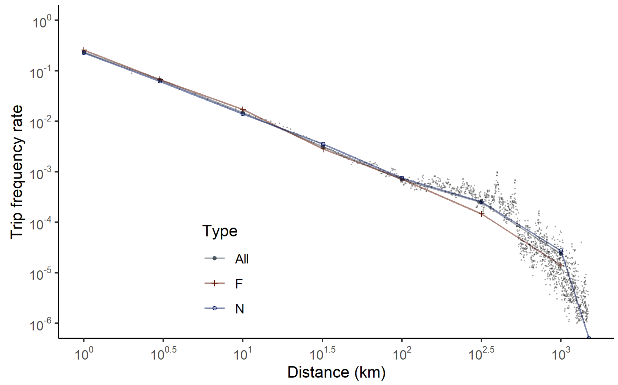

Figure 5: Distribution of distance-to-home of visited places. The gray points represent the frequency rates of all individuals across 1,500 distance groups. The larger points and lines depict results from selected distance groups, accentuating data trends and contrasting the differences between foreign-born and native-born.

In this simulation, we randomly shift the non-home stays to a POI within its own category or a similar kind that is d m away from the individual's home. This distance d is sampled from the average distance-to-home distribution (the gray points in Figure 5). Specifically, this shifted stay is randomly selected from the POIs between the search radiuses of $d + 30$ m and $d - 30$ m around the individual's home location. POI data and how the random shifting is implemented are detailed in Section A.7.1 Point of interest data.

Similarly to the No-destination preference simulation, we simulate 50 times for each individual. At last, we repeat the same calculation described in Section 4.1.6 to get the simulated values of experienced segregation.

Data availability

The data supporting the findings of this study were purchased from PickWell and are subject to restrictions due to licensing and privacy considerations under the European General Data Protection Regulation. Consequently, these data are not publicly available, but are commercially available and may be requested for research use (`https://www.pickwell.co/`). Aggregated data to reproduce all results are publicly available at `https://github.com/MobiSegInsights/mobi-seg-se/tree/main/data`. Venue locations and categories can be retrieved from OpenStreetMap (`https://download.geofabrik.de`). Census data (DeSO zones and their statistics) were collected from Statistics Sweden (`https://www.scb.se/`) that is publicly available. Census data (population grids) were collected from the Swedish University of Agricultural Sciences (`https://maps.slu.se/`) and were restricted to use only by individuals associated with Swedish research institutes. GTFS data were collected from Samtrafiken API (`https://samtrafiken.se/`) that are publicly available. All data were utilized in accordance with the terms of service specified by their respective provider.

We strictly adhered to the guidelines set forth by the Chalmers Institutional Review Board (IRB) in accordance with the Swedish Act (2003:460) concerning the ethical review of research involving humans, as well as the General Data Protection Regulation 2016/679 (GDPR). Due to the nature of the data analysed, the study was exempt from ethical review under the Swedish Ethical Review Act (2003:460).

Code availability

Python (version 3.10) code and R (version 4.0.2) code were used to analyse and visualize the data. The stays have been detected via infostop (version 0.1.11) and pyspark (version 3.5.1). Code to reproduce our results is publicly available on GitHub https://github.com/MobiSegInsights/mobi-seg-se.

References

Sako Musterd, Szymon Marcińczak, Maarten Van Ham, and Tiit Tammaru. Socioeconomic segregation in european capital cities. increasing separation between poor and rich. *Urban geography*, 38(7):1062–1083, 2017.

Raj Chetty, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. Social capital i: measurement and associations with economic mobility. *Nature*, 608(7921):108–121, 2022.

Sandro Sousa and Vincenzo Nicosia. Quantifying ethnic segregation in cities through random walks. *Nature Communications*, 13(1):5809, 2022.

John Iceland and Kyle Anne Nelson. The residential segregation of mixed-nativity married couples. *Demography*, 47 (4):869–893, 2010.

Bo Malmberg, Eva Andersson, and John Östh. Segregation and urban unrest in sweden. *Urban geography*, 34(7): 1031–1046, 2013.

Benjamin F Jarvis, Robert D Mare, and Monica K Nordvik. Assortative mating, residential choice, and ethnic segregation. *Research in Social Stratification and Mobility*, page 100809, 2023.

Yuan Liao, Jorge Gil, Sonia Yeh, Rafael HM Pereira, and Laura Alessandretti. Socio-spatial segregation and human mobility: A review of empirical evidence. *arXiv preprint arXiv:2403.06641*, 2024.

Otis Dudley Duncan and Beverly Duncan. A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217, 1955.

Flávia F Feitosa, Gilberto Camara, Antônio Miguel Vieira Monteiro, Thomas Koschitzki, and Marcelino PS Silva. Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science*, 21 (3):299–323, 2007.

Joana Barros and Flavia F Feitosa. Uneven geographies: Exploring the sensitivity of spatial indices of residential segregation. *Environment and Planning B: Urban Analytics and City Science*, 45(6):1073–1089, 2018.

Mei-Po Kwan. Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and giscience. *Annals of the Association of American Geographers*, 103(5):1078–1086, 2013.

Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications*, 12(1):4633, 2021.

Qing-Quan Li, Yang Yue, Qi-Li Gao, Chen Zhong, and Joana Barros. Towards a new paradigm for segregation measurement in an age of big data. *Urban Informatics*, 1(1):5, 2022.

Yoo Min Park and Mei-Po Kwan. Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation. *Computers, Environment and Urban Systems*, 71:98–108, 2018.

Sarah Garlick, Gemma Catney, Frances Darlington-Pollock, and Christopher D Lloyd. Is there greater ethnic mixing in residential or workplace spaces? *Journal of Ethnic and Migration Studies*, pages 1–21, 2022.

Luis Fuentes, Ricardo Truffello, and Mónica Flores. Impact of land use diversity on daytime social segregation patterns in santiago de chile. *Buildings*, 12(2):149, 2022.

Susan Athey, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt. Estimating experienced racial segregation in us cities using large-scale gps data. *Proceedings of the National Academy of Sciences*, 118(46):e2026160118, 2021.

Borja Moya-Gómez, Marcin Stępniak, Juan Carlos García-Palomares, Enrique Frías-Martínez, and Javier Gutiérrez. Exploring night and day socio-spatial segregation based on mobile phone data: The case of medellin (colombia). *Computers, Environment and Urban Systems*, 89:101675, 2021.

Eun Jin Shin. Spatial segregation of chinese immigrants in seoul, south korea, during the covid-19 pandemic: Evidence from population data derived from mobile phone signals. *The Social Science Journal*, pages 1–22, 2021.

Siiri Silm, Veronika Mooses, Anniki Puura, Anu Masso, Ago Tominga, and Erki Saluveer. The relationship between ethno-linguistic composition of social networks and activity space: A study using mobile phone data. *Social Inclusion*, 9(2):192–207, 2021.

Shi Xian, Zhixin Qi, and Ngai-ming Yip. Beyond home neighborhood: Mobility, activity and temporal variation of socio-spatial segregation. *Journal of transport geography*, 99:103304, 2022.

Yang Xu, Paolo Santi, and Carlo Ratti. Beyond distance decay: Discover homophily in spatially embedded social networks. *Annals of the American Association of Geographers*, 112(2):505–521, 2022.

Bruna Da Cunha Kronenberger and Renato TibiriÇÁ De Saboya. A configurational study of sociospatial segregation in the metropolitan region of florianopolis, brazil. In *Proceedings 11th International Space Syntax Symposium*, volume 11, pages 75–1, 2017.

Laura H Atuesta, J Eduardo Ibarra-Olivo, Nancy Lozano-Gracia, and Uwe Deichmann. Access to employment and property values in mexico. *Regional Science and Urban Economics*, 70:142–154, 2018.

Lina Hedman, Kati Kadarik, Roger Andersson, and John Östh. Daily mobility patterns: Reducing or reproducing inequalities and segregation? *Social Inclusion*, 9(2):208–221, 2021.

Rafiazka Millanida Hilman, Gerardo Iñiguez, and Márton Karsai. Socioeconomic biases in urban mixing patterns of us metropolitan areas. *EPJ data science*, 11(1):32, 2022.

Yanji Zhang, Liang Cai, Guangwen Song, Lin Liu, and Chunwu Zhu. From residential neighborhood to activity space: The effects of educational segregation on crime and their moderation by social context. *Annals of the American Association of Geographers*, 112(8):2393–2412, 2022.

Karl Vachuska. Racial segregation in everyday mobility patterns: Disentangling the effect of travel time. *Socius*, 9: 23780231231169261, 2023.

Timur Abbiasov, Cate Heine, Sadegh Sabouri, Arianna Salazar-Miranda, Paolo Santi, Edward Glaeser, and Carlo Ratti. The 15-minute city quantified using human mobility data. *Nature Human Behaviour*, pages 1–11, 2024.

OECD. *International migration outlook 2022*. OECD, 2022.

Yang Xu, Alexander Belyi, Iva Bojic, and Carlo Ratti. Human mobility and socioeconomic status: Analysis of singapore and boston. *Computers, Environment and Urban Systems*, 72:51–67, 2018.

Hamed Nilforoshan, Wenli Looi, Emma Pierson, Blanca Villanueva, Nic Fishman, Yiling Chen, John Sholar, Beth Redbird, David Grusky, and Jure Leskovec. Human mobility networks reveal increased segregation in large cities. *Nature*, pages 1–7, 2023.

John Östh, Ian Shuttleworth, and Thomas Niedomysl. Spatial and temporal patterns of economic segregation in sweden's metropolitan areas: A mobility approach. *Environment and Planning A: Economy and Space*, 50(4): 809–825, 2018.

Izhak Schnell and Nasreen Haj-Yahya. Arab integration in jewish-israeli social space: does commuting make a difference? *Urban Geography*, 35(7):1084–1104, 2014.

Xiang Zhou and Yuning Cheng. Between state and family: discussion on the segregation and integration of the daily living space within shanghai historic lane neighborhood. *Home Cultures*, 16(3):163–190, 2019.

Michael Dorman, Tal Svoray, and Itai Kloog. How does socio-economic and demographic dissimilarity determine physical and virtual segregation? *Journal of Spatial Information Science*, (21):177–202, 2020.

Xiao Huang, Yuhui Zhao, Siqin Wang, Xiao Li, Di Yang, Yu Feng, Yang Xu, Liao Zhu, and Biyu Chen. Unfolding community homophily in us metropolitans via human mobility. *Cities*, 129:103929, 2022.

Rafael HM Pereira, Tim Schwanen, and David Banister. Distributive justice and equity in transportation. *Transport reviews*, 37(2):170–191, 2017.

Ulf Aslak and Laura Alessandretti. Infostop: scalable stop-location detection in multi-user mobility data. *arXiv preprint arXiv:2003.14370*, 2020.

Ramaswamy Hariharan and Kentaro Toyama. Project lachesis: parsing and modeling location histories. In *International Conference on Geographic Information Science*, pages 106–124. Springer, 2004.

Statistikmyndigheten SCB. DeSO – Demografiska statistikområden, 2023. URL `https://www.scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden/`.

Douglas S Massey. Residential segregation and neighborhood conditions in us metropolitan areas. *America becoming: Racial trends and their consequences*, 1(1):391–434, 2001.

Official Statistics of Sweden. *Swedish National Travel survey (RVU Sweden) 2011—2016*, 2016. URL `https://www.trafa.se/en/travel-survey/travel-survey/`.

Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3):278–295, 2013.

Wendy Van de Kerckhove, Leyla Mohadjer, and Thomas Krenzke. A weight trimming approach to achieve a comparable increase to bias across countries in the programme for the international assessment of adult competencies. *JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association*, pages 655–666, 2014.

Google. GTFS static dataset, 2019. URL https://gtfs.org/reference/static.

Hao Wu, Paolo Avner, Genevieve Boisjoly, Carlos KV Braga, Ahmed El-Geneidy, Jie Huang, Tamara Kerzhner, Brendan Murphy, Michał A Niedzielski, Rafael HM Pereira, et al. Urban access across the globe: an international comparison of different transport modes. *npj Urban Sustainability*, 1(1):16, 2021.

Rafael HM Pereira, Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. r5r: rapid realistic routing on multimodal transport networks with r 5 in r. *Findings*, 2021.

Acknowledgements

This research is funded by the Swedish Research Council (Project Number 2022-06215).

Author contributions

Y.L. and L.A. conceptualized the study. Y.L., J.G., and R.H.M.P. processed the data. Y.L. conducted simulations and data analysis. All authors wrote the manuscript.

Funding

Open access funding is provided by Chalmers University of Technology.

Competing interests

The authors declare that there are no conflicts of interest.

Additional information

Correspondence and requests for materials should be addressed to Y.L.

A Data and processing

A.1 Mobility data

Mobile application data consists of GPS records collected through location-enabled applications installed on people's smartphones, with the form of (*id, lat, lon, time*). This study uses a dataset from a diverse set of mobile apps used by anonymized adult smartphone users in Sweden[3].

The dataset contains geolocations of moving (Figure A.1a) and being stationary (Figure A.1b) collected from cell towers, GPS, and Wi-Fi sources, depending on the geolocation source at the moment when specific phone applications being used. Therefore, this dataset features varying sampling frequencies and spatial resolutions. It also contains sampling bias, given that geolocations are collected only when the users use certain applications on their phones.

A.2 Stay detection

Infostop algorithm has four configurable parameters to identify stays, summarized in Table A.1. The two spatial parameters, r_1 and r_2, control the spatial accuracy of identified stays; the greater their values, the less accurate the identified stays. Too small values of these two parameters make the algorithm sensitive to GPS noise. To balance spatial accuracy and robustness against noise, we choose 30 m based on the raw data's distribution.

The minimum duration of a stay is 15 min because we want to approximate the social interactions by identifying individuals' stays for activities. A too-short stay is unlikely to trigger social interactions with those in the same place. While an overly long minimum duration of a stay may miss some activities, especially considering the data sparsity issue.

[3]http://www.pickwell.co/

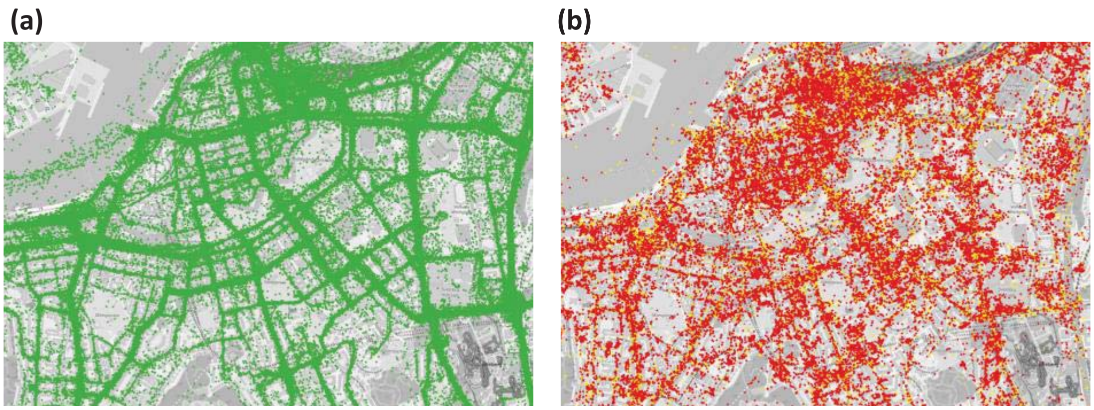

Figure A.1: Geolocations from mobile application data in the central Gothenburg. **(a)** Movement points. **(b)** Stationary points.

Table A.1: Infostop algorithm parameters and set values.

Parameter	Definition	Unit	Value
r_1	The maximum roaming distance allowed for two points within the same stay	meter	30
r_2	The typical distance between two stays at the same destination	meter	30
t_{min}	The minimum duration of a stay	min	15
t_{max}	The maximum time difference between two consecutive records to be considered within the same stay	hour	3

The last parameter, t_{max}, affects the maximum stay duration and how many records will be included in one stay. The raw geolocation data are passively collected only when people use mobile phone applications. Therefore, these geolocations are sparse in the time dimension and need careful exploration to determine a reasonable value for t_{max} to avoid creating artifacts in the detected stays. The artifacts originate from unobserved locations in the data between two consecutive geolocations close to each other in space. If we pick a big t_{max} to detect stays, these two close geolocations, interrupted by other movements in between but unobserved from the data, will be falsely merged into one stay.

To determine the t_{max} value for stay detection, we test 1 to 12 hours in 1 hour intervals and quantify its impact on the detected stays. In the lack of ground truth of individuals' whereabouts, we examine the resulting stays regarding their duration distribution, i.e., the share of stays with duration above t_{max}, and their underlying records, i.e., the percentage of stays with average geolocations per hour above 1. Figure A.2 shows the impact of t_{max} on the detected stays.

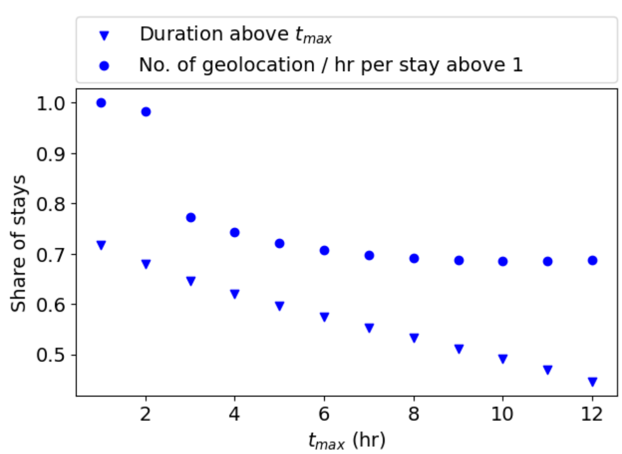

Figure A.2: Impact of t_{max} on the detected stays (a sample dataset).

We choose 3 hours for t_{max} to detect stays. This ensures sufficient geolocations underlying a stay (c.a. 80% stays above one geolocation per hour) and reasonable stay durations (35% stays below 3 hours). According to the Swedish National Travel Survey, 37% of activities are below 3 hours [Official Statistics of Sweden, 2016].

A.2.1 Data filtering

The individuals meeting the below criteria are selected for further analysis.

- The identified home has a corresponding grid in Sweden's population grids and DeSO zones. In other words, the individuals are presumed to live in Sweden.
- One should have at least three nights at home.
- The identified stays above 12 hours are removed.
- The individuals should have more than seven active days, and the number of unique locations should be more than two.

A.2.2 Data description

After processing the mobility data, we have 30,454,903 stays from 322,919 individual devices for further analysis. Their stays have the characteristics shown in Table A.2.

Table A.2: Descriptive stay statistics. An active day is when at least one stay is detected.

Attribute	min	25%	50%	75%	max
No. of unique locations	3	4	7	18	594
No. of active days	8	21	37	68	215
No. of stays	5	27	54	117	1882
Median stay duration (min)	20	180	183	194	4207

A.3 Census data

DeSO - Demographic statistical areas are a national subdivision defined by Statistics Sweden, following Sweden's county and municipal boundaries [Statistikmyndigheten SCB, 2023]. There are 5,984 DeSO zones, each with a unique 9-digit code, and the first three digits indicate which county it belongs to. The fifth position is a letter: A, B, or C, which groups DeSO zones into three different categories, where in this study, A and B stand for Rural/Suburban and C for Urban.

DeSO zones have varying sizes; however, these zones are designed to reflect the population distribution. Urban areas with high population densities have a high spatial resolution, i.e., they are represented by many small DeSO zones. Therefore, this study uses DeSO zones as the spatial unit to calculate visiting and experienced segregation metrics.

The statistics agency in Sweden (SCB) also provides grid-level information based on the total population register. We collect these population grids from the Swedish University of Agricultural Sciences[4]. To protect privacy, there are two spatial units: 250 m x 250 m (small) and 1000 m x 1000 m (large). The used population grids have 178,920 cells, where 57% are large, and 43% are small grid cells. The population size in each grid ranges between 0 and 7,844, with a median value of 13. We keep these 7% zero-population grid cells because they have non-zero daytime jobs used for transport access computation.

A.4 Analysis zones

To balance data sufficiency and analysis accuracy, the analysis zones for this study combine the census zones and the H3 zoning system[5]. Within each census zone, we create its children hexagons under H3, a geospatial indexing system that partitions the world into hexagonal cells. The size/resolution of the hexagons within each census zone depends on the census zone's area, as detailed in Table A.3.

[4] https://maps.slu.se/

[5] Hexagonal hierarchical geospatial indexing system. https://h3geo.org/

Table A.3: Analysis zones created combining parental census zones and children hexagons. There are 70,830 analysis zones, contrasting with the 5,984 census zones.

Parental census zone area (km^2)	No. of census zones	Children H3 zone resolution	Average area of analysis zones (km^2)
<0.5	1509	None[1]	0.2
0.5-3.5	1817	9	0.1
3.5-15	1424	8	0.7
15-100	440	7	5.2
100-720	683	6	36.1
>720	111	5	252.9

A.5 Sampling debiasing

Unlike travel surveys or constant GPS tracking, mobile phone application data only gives the population's geolocation when they use certain applications. This passive data collection results in imbalanced knowledge of where the individuals are (Figure A.3). To reduce such bias, this step assigns a weight to each detected stay individually, considering the overall temporal patterns of each device.

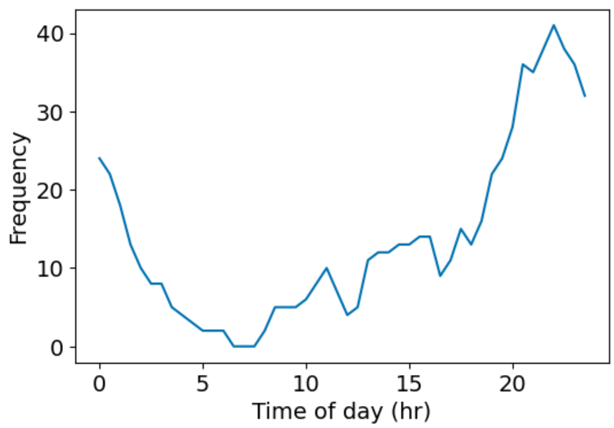

Figure A.3: Temporal distribution of detected stays of an exemplary individual.

For a given individual device, we first divide 24 hours into 48 groups with evenly placed half-hour intervals and then calculate the number of observed locations in each interval (f_i, i=1,2,...,48). Next, each time interval is assigned a weight ($w_i = 1/f_i$, i=1,2,...,48). Finally, we calculate how many intervals a detected stay spans over, e.g., a stay starting from 8 am till 9 am spans over the groups $17, 18$. And its weight is defined as $W = w_{17} + w_{18}$. The stay weights mimic the even sampling of a person's whereabouts, leading to a statistically smooth timeline.

A.6 Population debiasing

After identifying the individuals' homes, we know how many live in each DeSO zone. The Spearman correlation between the number of devices and the population size has a coefficient of 0.47 ($p < 0.001$), indicating mobile phone data's magnitude-wise representation of the actual population.

Figure A.4 shows the distributions of individual device numbers compared to the actual population sizes. Figure A.4a confirms the medium-level correlation between the individual device number and the population size in DeSO zones. Figure A.4b shows to what extent DeSO zones' populations are represented in the mobile phone data. The share of the population represented ranges between 0.19% and 93%, except for six outlier zones with more devices than the actual population.

We use inverse probability weighting (IPW) to assign each device a weight [Seaman and White, 2013], i.e., the reverse of the phone users' count ratio to the DeSO zone's actual population size (W_p). This individual weight, W_p, has extreme values in some DeSO zones where only a few devices are included. Weight trimming technique is applied [Van de

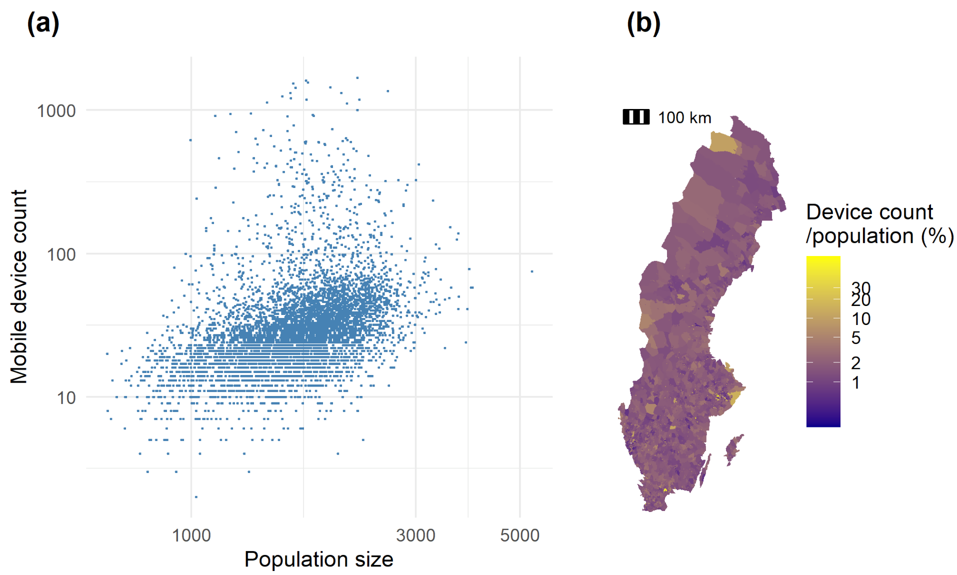

Figure A.4: Home distribution. **(a)** Number of devices vs. actual population size by DeSO zone. **(b)** Share of devices compared to the actual population by DeSO zone.

Kerckhove et al., 2014], where any weight above the cut-point weight (W_0) is set to W_0. Equation 2 determines this cut-point weight value.

$$W_0 = 3.5\sqrt{1 + \texttt{CV}^2\left(\mathbf{W_p}\right)} \times \texttt{Med}\left(\mathbf{W_p}\right) \tag{2}$$

where CV is the coefficient of variance and Med is the median value.

A.7 Transportation data

Points of interest and road networks are retrieved from OpenStreetMap archives of Sweden[6].

A.7.1 Point of interest data

Each point of interest (POI) has a class and subclass defined by OSM contributors. For example, a geolocation is labeled as “class=amenity” and “subclass=toilets.” There are over a thousand subclasses in the original Swedish POI dataset. For the ease of simulation, this study cleans the POI dataset and groups them into broader but smaller number of categories.

We first use GPT-4 to learn 24 preliminary categories based on the POIs’ subclass labels of Sweden. Then we manually check the validity of this preliminary category assigned to each POI’s subclass. We also exclude those POIs such as emergency point, toilet, etc. where people spend limited time and the social interactions are unlikely to happen. In addition, we merge some classes: tourism, historic = tourism, leisure, sport = leisure, craft, office = office, and distinguish POIs of the same sublcass but different classes of amenity or shop. At last, there are 33 categories of POIs where a stands for amenity and s stands for shop: Artisan Workshops, Automotive Services (a), Automotive Services (s), Craft, Education (a), Education (s), Entertainment (s), Fashion and Accessories (s), Financial Services (a), Financial Services (s), Food and Drink (a), Food and Drink (s), Groceries and Food (a), Groceries and Food (s), Health and Beauty (a), Health and Beauty (s), Healthcare (a), Healthcare (s), Home and Living, Leisure, Office, Office (s), Outdoor Recreation (a), Outdoor Recreation (s), Recreation (a), Recreation (s), Religious Places (a), Shop, Sports and Activities (a), Sports and Activities (s), Tourism, Transportation (a), Transportation (s).

Randomly shifting original stay records to a nearby similar kind of POI follows a few steps. The candidate nearby POIs shall belong to the same category. If none is found, then we do not distinguish amenity and shop categories, Office and Craft, or any categories belonging to any shops and the overall Shop category. The shifted POI is randomly selected from the POI candidate set, if any.

[6] https://download.geofabrik.de/europe.html

A.7.2 Public transit data

General Transit Feed Specification (GTFS) is an open data standard for public transport timetables proposed by Google. A GTFS static dataset [Google, 2019] is a collection of text files consisting of all the information required to reproduce a transit agency's schedule, including the locations of stops and timing of all routes and vehicle trips. This project collects the up-to-date GTFS data of Sweden from Samtrafiken AB[7]. The GTFS data are processed to remove the invalid shapes, extract the geolocations of all the transit stops, and calculate access to jobs by public transport.

A.8 Transport access

Transport access quantifies how easily one can reach various destinations. In this study, we consider taking public transit from the residence grid. We use a cumulative-opportunity indicator [Wu et al., 2021] to measure the number of jobs accessible within 30 min travel by public transport (A_t). The other relevant parameters, such as walking speed, are the default settings of *accessibility* function in *r5r*, an open-source transportation analysis package [Pereira et al., 2021].

B Modeling segregation by birth background

B.1 Adjusted ICE vs. its original form

The original formula of ICE is shown below (Equation 3).

$$ICE_j = \frac{N_j - F_j}{N_j + F_j + O_j} \tag{3}$$

For the comparison, Figure B.1 shows the distribution of DeSO zones regarding the two ICE metrics using Equations 3 and 1.

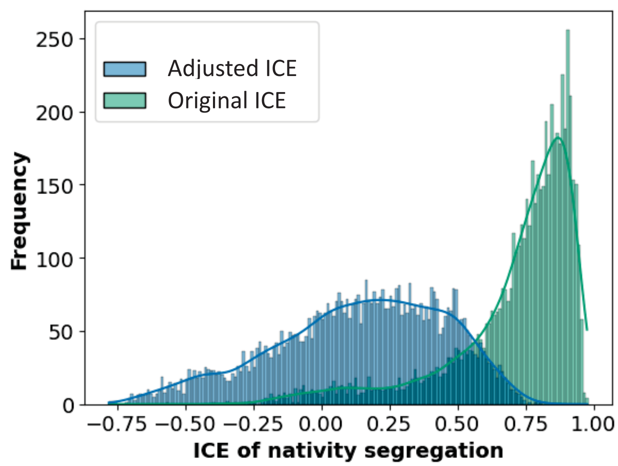

Figure B.1: Residential segregation by birth background: adjusted vs. original.

B.2 Activity patterns

In our dataset, foreign-born and native-born individuals show limited differences in activity participation based on visits to various types of POIs (Figure B.2). The differences measured by Cohen's d between Group F and Group N across all nine categories of POIs indicate negligible disparities, ranging from -0.08 to 0.14.

B.3 Individual attributes

Figure B.3 shows the distributions of individual attributes. Compared with residential segregation, the experienced and simulated segregation decrease collectively. Car ownership is predominantly skewed toward smaller values with a long tail. The number of accessible jobs by transit is generally smaller than that by car. At the same time, a few regions located at the very center of the cities have an exceptionally high density of transit services.

[7] https://samtrafiken.se/

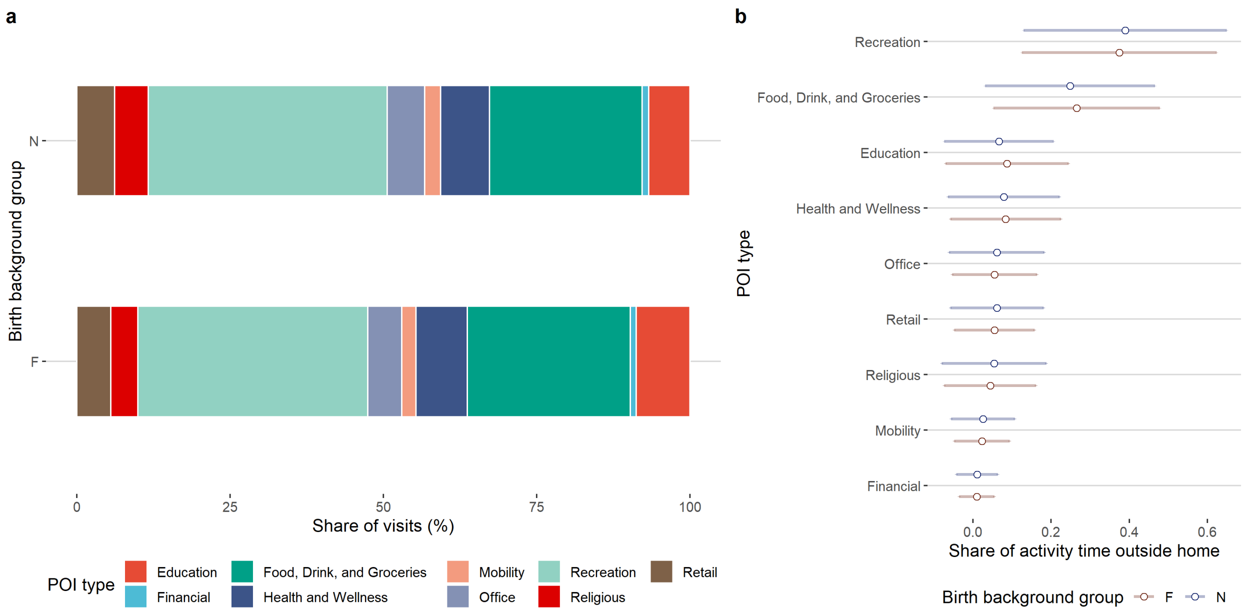

Figure B.2: Activity patterns. **a**, Share of visits by birth background group. **b**, Share of activity time by POI type and birth background group. Lines show the ranges between the values of weighted mean ± standard deviation.

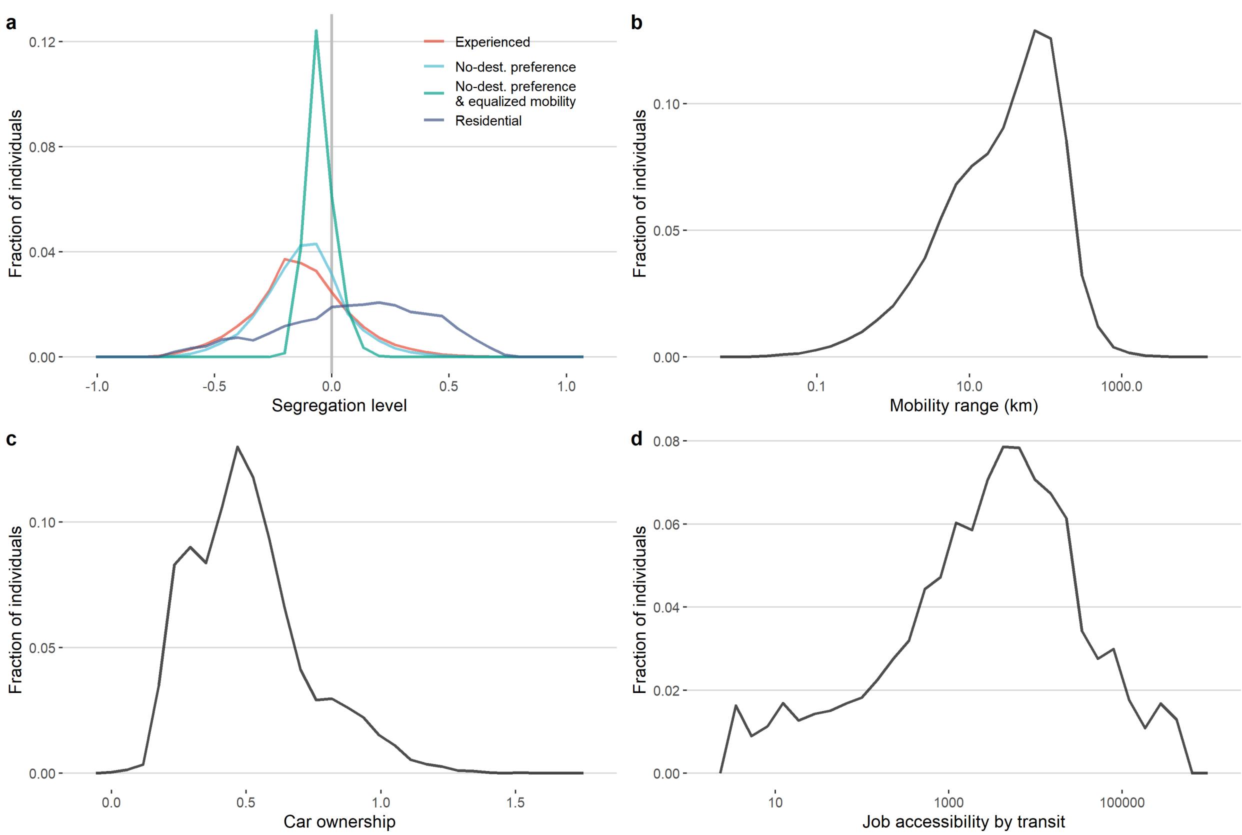

Figure B.3: Weighted individual distribution of key attributes. **a**, Experienced, residential, and simulated segregation levels. **b**, Mobility range. **c**, Car ownership per capita. **d**, Job accessibility by transit.

B.4 Impact of mobility range

Individuals in foreign-born segregated areas exhibit distinct characteristics regarding car ownership, transport access, and their experienced segregation, in contrast to those living in native-born concentrated areas (Table B.1). Over 50% of individuals in Group N have a high car ownership, while in Group F, the majority have a low to medium car ownership. Car ownership negatively correlates with job accessibility by transit. Group F has lower public transport access than Group N, except for those with high car ownership. This is especially drastic for those with low levels of car ownership. These individuals living in native-born segregated areas have almost nine times more accessible jobs by taking transit than their foreign-born counterparts. This may be related to the distinction between Groups N and F regarding where they live, primary travel mode, economic circumstances, and individual preferences.

Table B.1: Residents of the foreign-born and native-born segregated areas: average values of key statistics in segregation and transport access considering car ownership levels. Car ownership is divided into three groups representing the lowest 25%, the middle 50%, and the highest 25%: Low (< 0.28), Medium (0.28–0.72), and High ($>$0.72). F = Individuals living in foreign-born segregated areas, N = Individuals living in native-born segregated areas.

Group	Car ownership	Share in Group (%)	Job access by transit ($\times 10^3$)	Experienced segregation
	Low	38.4	17.8	-0.32
F	Medium	58.8	6.2	-0.29
	High	2.8	1.3	-0.21
	Low	2.1	203.4	-0.05
N	Medium	39.3	3.2	-0.06
	High	58.6	0.1	-0.05

B.5 Statistical test results

We present a summary of segregation metrics and group exposure statistics, including statistical test results. All median values are calculated using a weighted bootstrap method with 1000 repetitions, and the error is represented by the standard deviation of the bootstrap estimates. The statistical tests are conducted using the Wilcoxon weighted test to determine significant deviations from the baselines: segregation (Tables B.2-B.3 and Figure B.4) and deviation from random mixing group exposure (Table B.4 and Figure B.5).

Table B.2: Residential segregation level (ICE_r) by group. Errors are computed as the median values' bootstrap standard deviation (1000 repetitions). * This is defined by 99% confidence interval around zero in the randomized mixing scenario. See details in Section 4.2.1.

Group	Median	Error	Segregation*
F	0.385	0.001	Yes
N	-0.380	0.002	Yes
M	0.029	0.001	No

Table B.3: Experienced segregation level (ICE_e) by group outside residential area. Errors are computed as the median values' bootstrap standard deviation (1000 repetitions). $^*p < 0.001$ The test compares the group ICE_e distributions with the random mixing thresholds defined in Section 4.2.1. If a group distribution is not greater than 0.2 and not smaller than -0.2, we consider it insignificant segregation (Segregation = No); otherwise, we have Segregation = Yes.

Group	Scenario	Median	Error	Segregation*
	Empirical	-0.295	0.001	Yes
F	No dest. Preference	-0.234	0.001	Yes
	Equalized mobility & no dest. preference	-0.066	0.0003	No
	Empirical	-0.054	0.001	No
N	No dest. Preference	-0.060	0.001	No
	Equalized mobility & no dest. preference	-0.049	0.0003	No
	Empirical	-0.146	0.001	No
M	No dest. Preference	-0.125	0.001	No
	Equalized mobility & no dest. preference	-0.062	0.0002	No

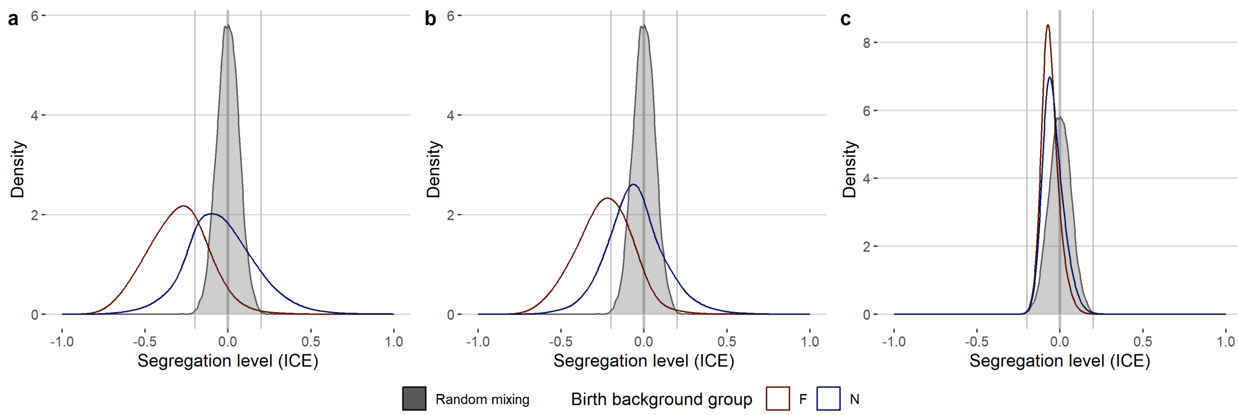

Figure B.4: Experienced segregation levels outside residential area: comparison with Random mixing. Vertical lines mark the thresholds, -0.2 and 0.2, for the statistical test to determine whether the distribution is significantly segregated. **a**, Experienced. **b**, No dest. Preference. **c**, Equalized mobility & no dest. preference.

Table B.4: Group exposure outside residential area ($X \to X$), share by group (%). Errors are computed as the median values' bootstrap standard deviation (1000 repetitions). * Deviation from random mixing, $p < 0.001$. The test compares the group exposure distributions with the random mixing thresholds defined in Section 4.2.1. For $X \to D$ and $X \to M$, if a group distribution is not greater than 19% and not smaller than -19%, we consider it an insignificant deviation from homogeneous exposure (Deviation = No); otherwise, we have Deviation = Yes. For $X \to F$, the criterion is not greater than 13% and not smaller than -13% for being insignificant.

Exposure	Scenario	Median	Error	Deviation*
	Empirical	30.0	0.17	Yes
$F \to F$	No dest. Preference	6.9	0.07	No
	Equalized mobility & no dest. preference	4.3	0.01	No
	Empirical	-23.3	0.08	Yes
$F \to N$	No dest. Preference	-14.3	0.08	No
	Equalized mobility & no dest. preference	-7.3	0.01	No
	Empirical	17.7	0.14	No
$N \to N$	No dest. Preference	0.6	0.10	No
	Equalized mobility & no dest. preference	-1.8	0.02	No
	Empirical	-8.7	0.05	No
$N \to F$	No dest. Preference	-3.1	0.02	No
	Equalized mobility & no dest. preference	0.6	0.01	No

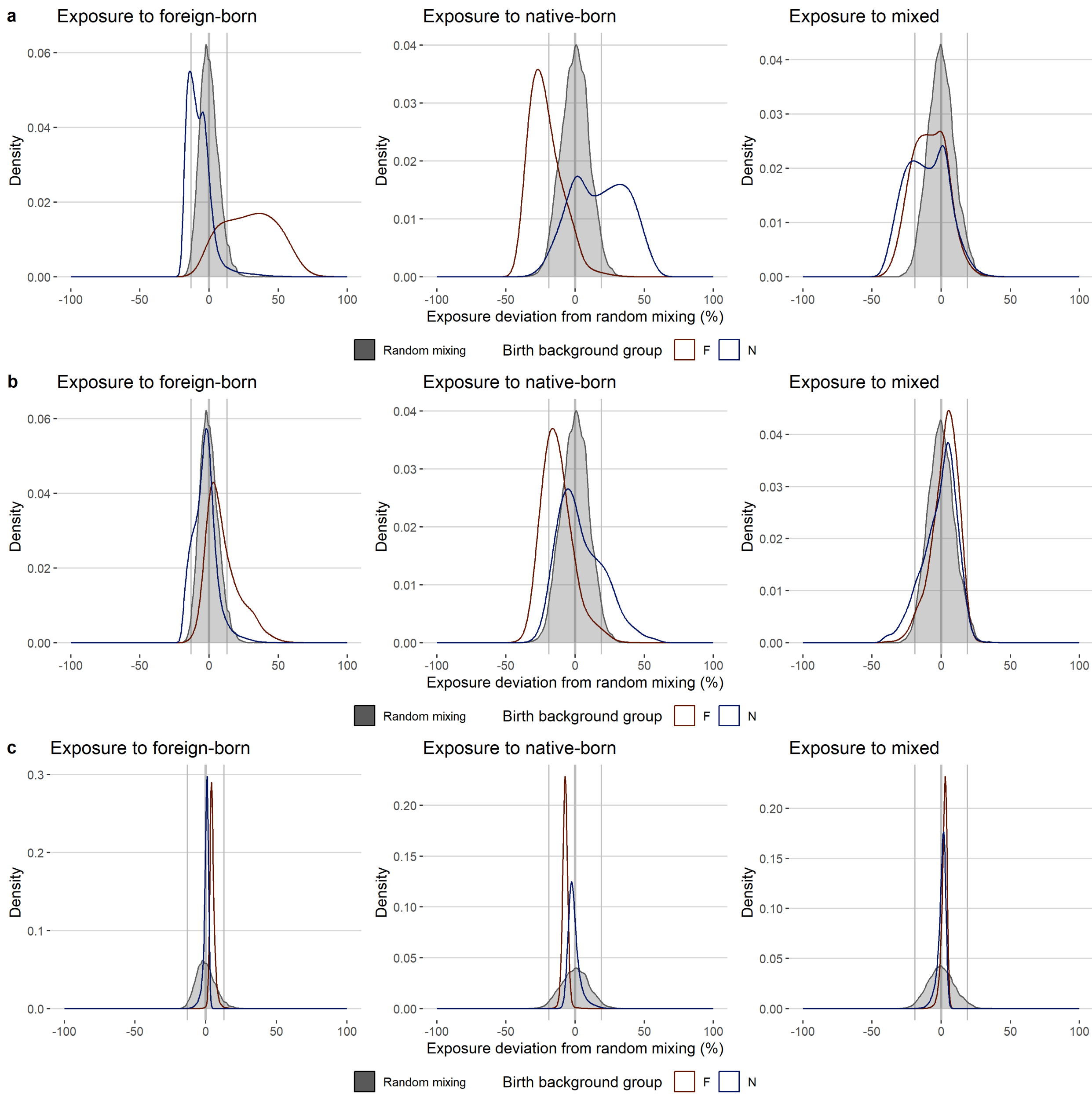

Figure B.5: Group exposure outside residential area: comparison with Random mixing. Vertical lines mark the thresholds, -13 and 13 for F, and -19 and 19 for N and M, for the statistical test to determine whether the distribution deviates significantly from Random mixing. **a**, Experienced. **b**, No dest. Preference. **c**, Equalized mobility & no dest. preference.